




Improved search engine

Publication number: CN1299488
Publication date: 2001-06-13
Inventor: RYAN GRANT J (US); RYAN SHAUN W (US); RYAN CRAIG M (US)
Applicant: NBEI NEW ZEALAND CO LTD (US)
Classification:
- international: **G06F17/30; G06F17/30; (IPC1-7): G06F17/30**
- european: G06F17/30T; G06F17/30W1
Application number: CN19998005701 19990316
Priority number(s): US19980078199P 19980316; US19980115802 19980715

Also published as:

 WO9948028 (A3)
 WO9948028 (A2)
 EP1072002 (A3)
 EP1072002 (A2)
 US6421675 (B1)

more >>

[Report a data error here](#)

Abstract not available for CN1299488

Abstract of corresponding document: **WO9948028**

The present invention provides for a method of updating an internet search engine database with the results of a user's selection of specific web page listings from the general web page listing provided to the user as a result of his initial keyword search entry. By updating the database with the selections of many different users, the database can be updated to prioritize those web listings that have been selected the most with respect to a given keyword, and thereby presenting first the most popular web page listings in a subsequent search using the same keyword search entry.

Data supplied from the **esp@cenet** database - Worldwide

Best Available Copy

[12] 发明专利申请公开说明书

[21] 申请号 99805701.0

[43] 公开日 2001 年 6 月 13 日

[11] 公开号 CN 1299488A

[22] 申请日 1999.3.16 [21] 申请号 99805701.0

[30] 优先权

[32] 1998.3.16 [33] US [31] 60/078,199

[32] 1998.7.15 [33] US [31] 09/115,802

[86] 国际申请 PCT/US99/05588 1999.3.16

[87] 国际公布 WO99/48028 英 1999.9.23

[85] 进入国家阶段日期 2000.10.31

[71] 申请人 NBCI 新西兰有限责任公司

地址 美国加利福尼亚

[72] 发明人 格兰特·J·瑞恩 肖恩·W·瑞恩

克雷格·M·瑞恩 韦恩·A·芒罗

黛尔·鲁宾逊

[74] 专利代理机构 中国国际贸易促进委员会专利商标事
务所

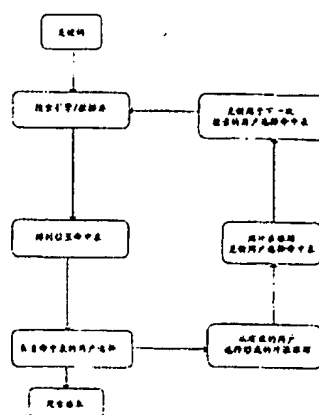
代理人 于 静

权利要求书 11 页 说明书 52 页 附图页数 27 页

[54] 发明名称 改进的搜索引擎

[57] 摘要

本发明提供一种利用用户对特定网页列表的选择结果更新因特网搜索引擎数据库的方法,该特定网页列表的选择结果来自提供给该用户作为其初始关键词搜索输入的结果的一般网页列表。通过用许多不同用户的选择来更新该数据库,可更新该数据库以便按重点排列相对给定关键词选择最多的那些网页列表,从而在以后使用相同关键词搜索输入的搜索中首先给出最流行的网页。



ISSN 1008-4274

一个相关联的那些数据项目产生所述多个列表。

12.根据权利要求9所述的方法，其中所述接收步骤接收第三所述用户群的标识；和

其中所述产生步骤仅从与所述第一，所述第二和所述第三识别用户群全部相关联的那些数据项目产生所述多个列表。

13.根据权利要求11所述的方法，其中所述接收步骤接收第二所述用户群的标识；和

其中所述产生步骤仅从与所述第一，所述第二或所述第三识别用户群中的一个相关联的那些数据项目产生所述多个列表。

14.根据权利要求1所述的方法，如果所述关键词是被允许的关键词则进一步包括确定步骤；和

其中如果所述关键词是被允许的关键词，则仅执行所述产生步骤。

15.根据权利要求1所述的方法，其中所述检测步骤仅在确定的时间间隔期间从每个用户网点检测每个关键词一次。

16.根据权利要求15所述的方法，其中检测每个关键词一次的所述步骤包括步骤：

将一个标识符与每个用户网点相关联；和

使用所述标识符跟踪已从所述多个用户网点中的每一个输入的关键词。

17.根据权利要求15所述的方法，其中所述产生步骤使用与确定的所述多个列表中的每个关键词相关联的历史系数。

18.根据权利要求16所述的方法，其中所述历史系数是小于或等于1且大于或等于0的数值。

19.根据权利要求1所述的方法，其中与每个所述数据项目相关联的是更新数据；和

其中所述产生步骤根据与最新更新的所述关键词对应的数据项目和最流行的数据项目来产生所述多个列表。

20.根据权利要求1所述的方法，其中与每个数据项目相关联的是

新近的加权系数X和老加权系数Y；和

其中所述产生步骤根据按使用所述新近的加权系数X和所述老加权系数Y确定的在流行性中增加最快的数据项目，和最流行的数据项目来产生所述多个列表。

21.根据权利要求1所述的方法，进一步包括步骤：

在接收所述关键词的所述步骤之前，接收识别用户的密码；和

其中所述产生步骤针对接收的所述关键词，和最新流行的数据项目产生与所述用户相关联的用户专用列表的组合作为所述多个列表，在与所述用户相关联的所述检测步骤的较早步骤中已检测到所述用户专用列表。

22.根据权利要求1所述的方法，其中与每个数据项目相关联的是新近的加权系数X和老加权系数Y；和

其中所述产生步骤根据按使用所述新近的加权系数X和所述老加权系数Y确定的在流行性中增加最快的数据项目，和与最新更新的所述关键词对应的数据项目来产生所述多个列表。

23.根据权利要求1所述的方法，进一步包括步骤：

在接收所述关键词的所述步骤之前，接收识别用户的密码；和

其中所述产生步骤针对接收的所述关键词，并依据与最新更新的所述关键词对应的数据项目产生与所述用户相关联的用户专用列表的组合作为所述多个列表，在与所述用户相关联的所述检测步骤的较早步骤中已检测到所述用户专用列表。

24.根据权利要求1所述的方法，进一步包括步骤：

在接收所述关键词的所述步骤之前，接收识别用户的密码；和

其中所述产生步骤针对接收的所述关键词，并根据使用所述最近的加权系数X和所述老加权系数Y确定的流行性增加最快的数据项目产生与所述用户相关联的用户专用列表的组合作为所述多个列表，在与所述用户相关联的所述检测步骤的较早步骤中已检测到所述用户专用列表。

25.根据权利要求1所述的方法，其中与每个所述数据项目相关联

在所述更新步骤之前，接收识别用户的密码；

使用所述密码确定所述用户是否是所选择的用户群中的一个；

和

其中如果所述用户不是所述选择的用户群中的一个，则不进行所述更新步骤。

44.根据权利要求43所述的方法，其中如果所述用户是所述选择的用户群中的一个并且在所述选择的数据项目中度过预定的时间周期则产生所述用户跟踪。

45.根据权利要求28所述的方法，其中在确定的时间间隔期间所述更新步骤仅针对每个用户更新每个关键词一次。

46.根据权利要求45所述的方法，其中更新每个关键词一次的所述步骤包括步骤：

使标识符与每个用户相关联；和

使用所述标识符跟踪每个用户已输入的关键词。

47.在具有多个开发者网点的计算机网络中，一种确定与从服务器计算机发送的其它信息一起提供给用户网点的内容的方法，包括步骤：

从所述多个所述开发者网点中的每一个获得内容列表，每个所述内容列表包括内容，开发者标识符，关键词，和关键词选择系数；

从所述获得的关键词确定对不同内容列表相同的特定关键词；

和

在确定向所述用户网点之一发送与所述内容列表之一相关联的特定内容，而不是与另一个所述内容列表相关联的其它内容的过程中使用该关键词选择系数。

48.根据权利要求47所述的方法，进一步包括获得每个内容列表的简介和简介选择系数；

从所述获得的简介确定对不同内容列表相同的特定简介；和

在确定向所述用户网点之一发送与所述内容列表之一相关联的特定内容，而不是与另一个所述内容列表相关联的所述其它内容的

过程中使用该简介选择系数。

49.一种确定向多个用户电显示多个内容提供者的内容的百分比数值的方法，包括步骤：

从所述多个内容提供者中的每一个接收报价数值和相关联的特定关键词，所述报价数值对应于每个所述内容提供者与用于显示一部分内容的特定关键词相关联的值；

针对相同关键词使报价数值相关，以便确定将每部分内容电显示给所述多个用户的百分比数值；和

部分地根据该百分比数值来显示每个所述部分的内容。

50.根据权利要求49所述的方法，其中所述百分比数值以可接收所述特定关键词的时间周期为基础。

51.根据权利要求49所述的方法，其中所述百分比数值以所述特定关键词的发送数量为基础。

52.根据权利要求49所述的方法，进一步包括步骤：

接收所述特定关键词的附加报价数值；

使报价数值与该附加报价数值再相关，以便确定将每个内容提供者的内容电显示给所述多个用户的更新的百分比数值；和

部分地根据更新的百分比数值显示每部分内容。

53.根据权利要求52所述的方法，其中所述更新的百分比数值以可接收所述特定关键词的时间周期为基础。

54.根据权利要求52所述的方法，其中所述更新的百分比数值以所述特定关键词的发送数量为基础。

55.根据权利要求49所述的方法，进一步包括步骤：

接收所述特定关键词的修改报价数值来代替所述报价数值之一；

使报价数值与该修改报价数值再相关，以便确定将每个内容提供者的内容电显示给所述多个用户的更新的百分比数值；和

部分地根据更新的百分比数值显示每部分内容。

56.根据权利要求55所述的方法，其中所述更新的百分比数值以

可接收所述特定关键词的时间周期为基础。

57.根据权利要求55所述的方法，其中所述更新的百分比数值以所述特定关键词的发送数量为基础。

58.一种确定向多个用户电显示多个内容提供者的内容的百分比数值的方法，包括步骤：

从所述多个内容提供者中的每一个接收报价数值和相关联的简介，所述报价数值对应于每个所述内容提供者与用于显示一部分内容的特定关键词相关联的值；

针对相同简介使所有报价数值相关，以便确定将每部分内容电显示给所述多个用户的百分比数值；和

根据该百分比数值来显示每个所述部分的内容。

59.根据权利要求58所述的方法，其中所述百分比数值以可接收所述简介的时间周期为基础。

60.根据权利要求58所述的方法，其中所述百分比数值以所述简介的发送数量为基础。

61.根据权利要求58所述的方法，进一步包括步骤：

接收所述简介的附加报价数值；

使报价数值与该附加报价数值再相关，以便确定将每个内容提供者的内容电显示给所述多个用户的更新的百分比数值；和

根据更新的百分比数值显示每部分内容。

62.根据权利要求61所述的方法，其中所述更新的百分比数值以可接收所述简介的时间周期为基础。

63.根据权利要求61所述的方法，其中所述更新的百分比数值以所述简介的发送数量为基础。

64.一种确定从内容提供者提供的多个不同的内容中向一个用户电显示哪一个内容的方法，包括步骤：

从所述多个内容提供者中的每一个接收报价数值，所述报价数值对应于用于显示每个所述内容提供者与一个用户群的至少一个关键词或一个简介相关联的一部分内容的值；

针对每个不同的报价数值将所有报价数值与不同的关键词和简介相关，以便能够确定对于每个不同的关键词和简介来说哪一个内容具有最高百分比数值；和

根据该用户的一个关键词和该简介向一个用户发送具有最高对应报价数值的一部分内容。

65.在具有多个用户网点的计算机网络中，一种对服务器计算机的数据库中存储的多个关键词的相对重要性加权的方法，包括步骤：

在所述服务器计算机从用户网点接收初始的关键词；

在所述服务器计算机产生与所述初始的关键词对应的多个有关关键词；

从所述服务器计算机向所述用户网点之一发送所述多个有关关键词；

在所述服务器计算机检测所述用户网点选择了所述多个有关关键词中的那一个；和

更新所述数据库，以便对所述未选择的关键词与所述初始关键词的关系的加权比所述选择的关键词与所述初始关键词的关系的加权大。

66.根据权利要求65所述的方法，其中存在将所述多个网点的所述用户分组的多个简介，其中更新步骤对所述选择的关键词的所述多个简介的不同简介进行不同的加权。

67.根据权利要求65所述的方法，其中当产生该多个相关的关键词时，所述产生步骤还考虑与提交初始关键词的用户网点相关联的简介。

68.根据权利要求65所述的方法，其中存储该多个相关关键词作为分级目录，以使更经常使用的关键词的等级较高。

69.根据权利要求65所述的方法，其中从所述初始关键词的错误拼写中部分地产生该多个相关关键词。

70.根据权利要求65所述的方法，其中从包含所述初始关键词的

短语中部分地产生该多个相关关键词。

71.根据权利要求65所述的方法，其中使用来自特定用户的过去喜好的信息部分地产生该多个相关关键词。

72.根据权利要求65所述的方法，其中从最新更新的关键列表和最流行的关键词部分地产生该多个相关关键词。

73.根据权利要求65所述的方法，其中多个用户群与每个关键词相关联；和

其中接收所述关键词的所述步骤还接收第一所述用户群的标识；和

其中所述产生步骤仅从与所述第一识别的用户群相关联的那些关键词产生所述多个相关关键词。

74.根据权利要求73所述的方法，其中所述接收步骤接收第二所述用户群的标识；和

其中所述产生步骤仅从与所述第一和所述第二识别的用户群二者相关联的那些关键词产生所述多个相关关键词。

说明书

改进的搜索引擎

本发明涉及允许增强的数据库搜索，特别是供互联网搜索引擎使用的方法和装置。

对获得相关信息并筛选不想要/不感兴趣的信息的有效和实用装置的需要正在增长，特别是由于互联网的创立。由于全球网站的数量在指数级地增长和其中包含的绝对信息量特别加剧了该需求。在检索因特网上可供使用的信息的尝试中，已经产生大量软件搜索引擎，用户借助这些搜索引擎在其个人计算机上从键盘输入由适当的关键词构成的搜索命令。搜索命令发送到服务器计算机，该服务器计算机具有与该服务器计算机有关的搜索引擎。搜索引擎接收搜索命令，然后用其通过网址的数据库和网点上存储的文本扫描这些关键词。此后，把扫描的结果从服务器计算机传回到用户计算机并在用户计算机的屏幕上显示。

为使该搜索引擎被新网点所了解和更新其现有网点的记录，或是网点所有者本身通知搜索引擎或可通过“网络浏览器”获得该信息以更新在服务器计算机的数据库。网络浏览器是一种自动程序，该自动程序探索和记录网点的内容和其到其它网点的链接，从而在尝试检索所有当前网点的网点之间扩展。

该数据库结构和搜索该数据库的方法造成某些明显的困难。尽管有频繁删除，重新寻址，更新网点的事实等，因特网增长率在新网点的扫描中已导致大量的积压，因而留下了过时和/或误导信息的搜索引擎。虽然可构成网络浏览器以便根据其位置(标题，嵌入链接，地址等)按重点排列可能的关键词，然而，根据所使用的搜索引擎，仍需要扫描网点文本(经常涉及大部分或甚至所有网点文本)的实质部分。这样导致对搜索引擎的巨大的存储需求。此外，典型的关键词搜索可能带来极大量的资料，用户对其中的大部分不感兴趣。

用户通常根据网点的简要说明从该列表进行选择，并探索所选择的网点直到找出需要的信息。

这些结果是根据搜索引擎的具体标准排列的列表形式。这些标准可以从关键词在搜索到的文本中任何地方出现的次数排列，到对特定位置(如前面提到的)中使用的关键词给出加权的方法。当已使用多个关键词时，还根据应用的不同关键词的数量排列站点。所有这些排列系统的基本缺陷在于其客观性，它们是根据搜索引擎的编程标准，和特定类型的网点设计上设置的重要性确定的，而不用实际用户意见的任何度量标准。这确实可导致荒谬的情况，由此在为确保最常用的搜索引擎的有利访问率的尝试下，某些设计者根据上面提到的标准故意构成其网点，有损于该网点的存在，可读性和内容。

本发明的目的是通过开发操作人员的脑力来改善上面提到的常规搜索引擎的缺陷。

本发明另一个目的是提供一种新的搜索引擎，与现有的软件引擎相比，新的搜索引擎通过减少的系统存储和/或计算需求提高了效率，可使用性和效果。

本发明的再一个目的是与生成或更新其数据的指示一起提供搜索数据的流行性的各种指示。

为获得上面提到的本发明的优点，本发明中包括的一个实施例提供一种利用用户选择的具体网页的结果更新因特网搜索引擎数据库的方法，该具体网页是从作为用户的初始关键词搜索输入向用户提供的一般网页列表列出的。通过用许多不同用户的选择更新数据库，可更新该数据库，以便按重点排列已相对给出的关键词最多选择的那些网络列表，并由此在以后使用相同关键词搜索输入的搜索中首先给出最通用的网页列表。

在本发明的另一个实施例中，提供一种确定与从服务器计算机发送到用户网点的列表一起提供的内容的方法。在该实施例中，从多个不同开发者网点中的每一个获得一个内容列表。每个内容列表包括内容，开发者标识符，和关键词，以及关键词选择系数。此后，

从获得的对不同内容列表相同的关键词确定特定的关键词。对于该特定关键词，在确定何时向用户网点发送不同的内容列表时使用关键词选择系数，

在另一个实施例中，提供一种利用从向用户给出的相关的关键词列表获得的具体关键词的用户选择结果更新关键词表的方法。通过利用与该相同关键词有关的许多不同用户的选择来更新数据库，当以后输入相同的关键词时可首先提供和给出适当的关键词。

通过结合附图研究下面优选实施例的详细说明可理解本发明的这些和其它优点，其中：

图1表示本发明的整体特性；

图2说明对每个不同能力搜索的各种输入，说明提供根据本发明的引擎的输出；

图3A和3B说明根据本发明在形成向终端用户提供的搜索结果的过程中选择网页的处理示意图；

图4说明根据本发明用于不同网页搜索的数据集。

图5表示以前描述的各种数据集，和根据本发明导致提供建议的网页的列表的各种输入和作用；

图6说明根据本发明的流行搜索的实施；

图7说明根据本发明的热度减弱(hot off)新闻搜索的实施；

图8说明根据本发明的高跳跃(high-flyer)搜索的实施；

图9说明根据本发明的随机搜索的实施；

图10说明根据本发明的以前收藏页搜索的实施；

图11说明根据本发明的集中搜索的实施；

图12说明根据本发明的数据生成搜索的实施；

图13说明根据本发明的定制搜索的实施；

图14说明根据本发明基于用户群身份的搜索的实施；

图15说明根据本发明的关键词消除器特性；

图16说明根据本发明确定应将搜索结果用于形成累积冲浪者跟踪表的过程；

图17说明根据本发明的网页的主动建议。

图18说明根据本发明的网页的被动建议。

图19提供根据本发明的建议关键词的示意图。

图20说明根据本发明生成建议的关键词的数据集的方式。

图21说明根据本发明可生成建议的关键词列表的各种方式。

图22说明根据本发明如何将内容附加到网页列表。

图23说明根据本发明的各种内容数据集和占据它们的操作。

图24说明根据本发明的各种内容数据集和用于从它们选择数据的操作。

图25说明根据本发明的网页列表和其它内容数据。

图1A和1B说明了本发明的全部特性中的某些特性，下文将对其进一步详细描述。首先指出，在下面的描述中在各个附图和描述中使用相同编号的参考标号表示相同或相似的结构，动作或处理步骤。

本发明最好在每个计算机通常包含微处理器，存储器，和调制解调器，并且某些计算机包含显示器等的网络环境中实施，如所熟知的。如图1B所示，示出多个用户网点/计算机100A-100D，以及多个服务器计算机102A-B，开发者网点/计算机104A-B。可以理解，在典型的因特网中，如所说明的，不同的服务器计算机102可互连在一起。此外，虽然仅示出几个用户网点，开发者网点和服务器计算机，可以理解，可将几千台这样的计算机互连在一起。

虽然为根据所示的服务器计算机102操作的编码程序指令的序列实现本发明的应用写出了本发明的具体实施例，可以用其它形式来代替实现这些程序指令的特定序列，例如具有满足下面描述的应用的特殊指令的处理器。

正如下文所说明的，实现对用户透明的附加操作以便根据目前做出的搜索获得将来的搜索结果。如图所示，本发明具有各种能力，在图1A的并行流程中说明了每种能力，图1A说明了可同时的不同能力的概要。就其整个能力来说，启动框10给出三种：建议网页12，

建议关键词14, 和内容建议16.

为使用户根据本发明选择网页12, 有一个选择将要进行的搜索类型的步骤18. 此后, 在步骤20, 输入从各种来源中的一种获得的搜索输入并与步骤18选择的算法一起使用以确定搜索结果. 然后向用户显示该搜索的结果, 分别如图中标为步骤22, 24, 和26的显示网页的生成表, 显示被动建议的网页, 和显示主动建议的网页的步骤所示. 下文更详细地描述该能力, 以及其如何实现.

为出现用户可选择或不选择的关键词建议, 最好有选择将使用的关键词搜索算法类型的初始步骤28. 虽然许多系统可仅有这样一种算法, 如下文所述, 各种算法都是可能的. 一旦选择了关键词搜索算法, 随后的步骤30根据用户输入的关键词处理关键词数据的当前组以确定相关联的关键词. 然后在步骤30中向用户显示该操作的结果. 下文更详细地描述该能力, 以及其如何实现.

前面提到的网页和关键词选择能力适用于终端用户的直接利益. 本发明的间接适用于终端用户的利益的另一种新特性直接适用于广告商, 因为它允许满足根据各种标准实时地选取目标. 如下文更详细描述, 初始选择将在步骤34确定如何选择内容的内容提供算法. 随后在步骤36根据来自用户和内容提供者的输入来确定给出哪个内容. 此后, 为用户显示观看的广告, 同时显示关键词和/或网页.

虽然图1说明了根据本发明的某些综合特性, 如前面提到的, 本发明的许多有益特性是用户不能察觉的, 但对用户是透明的. 然而, 它们是很明显的, 以便全面解释如何实现本发明并在下文说明.

图2说明了根据本发明到搜索引擎的各种输入, 并针对不同能力说明了将要提供的输出. 从搜索引擎用户可能输入的数据包括:

- 关键词52 - 这是用户输入以查找网页列表的文字或短语
- 简介类型54 - 这些是与他们自身相关联的人群, 例如美国人, 男性, 医生等
- 用户ID56 - 这是选择用该搜索引擎登记的每个用户的唯一身

分。可通过网上“甜饼”(cookie)或登录进行

- 搜索类型58 - 可由搜索者主动选择以确定他们喜欢的(流行, 新的等)搜索结果的类型

- 日期时间60 - 这是搜索者使用该系统时被动记录的

- IP地址62 - 这是搜索者使用该系统时被动记录的

- 其它64 - 包括其它个人化的信息, 例如搜索定制偏爱, 网页建议的关键词等。该信息由用户主动输入, 然后每当用户(由用户ID识别)使用该搜索引擎时用来个人化搜索结果。

来自网页开发者的数据包括:

- URL66 - 这是他们希望提交的网页的URL地址

- 说明68 - 这是其网页上2-3行的信息说明

- 关键词70 - 这些是网页开发者喜欢与其网页关联的关键词

- 目标读者72 - 这些是网页开发者特别想作为目标的目标读者

(简介类型54)

- 日期时间74 - 这是每个网页开发者提交网页时被动记录的

来自内容提供者的数据包括:

报价76 - 这些是如后面描述的内容的\$报价

内容详情78 - 包括内容提供者的所有详情, 包括地址内容详情等

来自其它搜索引擎80的结果 - 这些是来自其它现有的搜索引擎的关键词搜索的结果。

其它搜索引擎10的输出是:

网页90的列表 - 可根据输入的数据在下文进一步描述的网页确定步骤82中产生网页列表;

内容关键词92 - 该搜索引擎向用户建议在下文进一步描述的关键词确定步骤84中试图产生的其它关键词; 和

内容94 - 搜索引擎发送出在下文进一步描述的确定内容步骤86中产生的选择的内容。

为便于参考和帮助理解, 参考本发明的优选实施例重申并在下

面详述以前和后面提到的数据集定义(与此适应,以表格形式给出用来生成定义的数据集的有关数据集的结构)。此后更详细解释其中的某些方面以便完全讲授如何做出和使用本发明。

位置: 多个唯一信息实体:

网页: 呈网页URL(通用参考定位器)地址形式的位置。

关键词: 在搜索引擎中输入的文字或短语

命中表: 作为关键词搜索结果的网页(URL地址)表。该命中表排列与关键词有关的网页的关系。该命中表总是具有与其相关联的关键词。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词(临时) · 与关键词匹配的数据库(永久) 	<ul style="list-style-type: none"> · 命中表-排列的网页命中表(临时)

永久数据集: 保持的长项(虽然其随时改变)

临时数据集: 仅在该搜索期间生成

冲浪者跟踪: 这是用户如何搜索的手段。它是他们搜索的,后来选择的URL的关键词的踪迹以及他们在那里度过多长时间,可从其计算用户(冲浪者)的网页的排列。关键词搜索之后它是他们找到最有效的网页的手段。使用所有冲浪者踪迹的组合来生成用户的选择命中表。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词(临时)来自初始搜索结果的用户选择(临时),即访问的网页(URL) · 每个URL的时间度过 · 用户的IP地址 	<ul style="list-style-type: none"> 冲浪者跟踪-用户找到的对每个关键词 有用的用户网页表(可以是永久或临时的)

用户选择的命中表：这是与每个关键词相关联的网页的半永久排列并表示找到的与关键词相关联的每个网页对因特网用户多么有用。用新冲浪者踪迹逐渐更新用户的选择的命中表。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 冲浪者踪迹(可以是永久或临时的) · 用户的选择的命中表(永久)* 	<ul style="list-style-type: none"> · 新用户的选择的命中表-排列的"普通"网页的命中表(永久)

· 最初的用户选择的命中表将是冲浪者踪迹

新网页表：这是从网页开发者提交的URL生成的新网页的列表。当万维网开发者更新网页时，他们可提交该网页地址，有关该网页的简要信息和有关的开发者决定的关键词表。然后将该网页放置在每个关键词新网页列表的顶部。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 有关网络地址和关键词的所有网页开发者信息 	<ul style="list-style-type: none"> · 新网页列表(永久)

内容提供者的列表：这是通常必须付费以给出有关该关键词的内容的内容提供者的列表(与每个关键词相关联)。所付价格取决于其它内容提供者的数量，他们所花费的量和搜索关键词的时间量。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词 · 内容提供者对内容地点的报价 	<ul style="list-style-type: none"> · 内容提供者表-与每个关键词相关联的内容的列表(永久)

高跳跃命中表：这是以最高速度增加流行性的(与每个关键词相关联的)网页的列表。它表示用户选择的命中表中的网页上升速度有多快，用它作为确保新出现的网页上升到用户选择的命中表顶部的手段。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 老用户的选择的命中列表-(临时) · 新用户的选择的列表-(永久) 	<ul style="list-style-type: none"> · 高跳跃命中表：流行性上升最快的网页的排列表

个人命中表：这是各个用户针对他们过去进行的每个关键词搜索已找到的最有用的网页的列表。它象每个单独用户的自动书签数据集。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词 · 单独的冲浪者踪迹-(永久) 	<ul style="list-style-type: none"> · 个人命中表：单独用户过去找到的有用的网页的排列表

集中搜索命中表：这可以是上面任何命中表的组合。可有许多不同的方式组合这些命中表。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 浏览器命中表(临时) · 用户选择的命中表(永久) · 广告商的列表(永久) · 新网页列表(永久) · 高跳跃列表(永久) · 个人命中表(永久) 	<ul style="list-style-type: none"> · 集中搜索命中表-(缺省)关键词搜索后显示给用户的网页的排列命中表。它可以是上面任何命中表的组合(永久)

浏览器关键词表：这是用户可以找到的有用的关键词建议表。这是通过把用户输入的关键词与其它用户已尝试的关键词和短语数据库匹配找到的。虽然它是关键词排列而不是网页排列，这也等同于浏览器命中表。进行这项工作的方法使用与拼写检验器仅针对短语进行的相似算法。还根据前面从用户关键词序列选择的URL建议关键词。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词(临时) · 使用的所有关键词的数据库(永久) 	<ul style="list-style-type: none"> · 用户可能想尝试的其它关键词的排列命中表(临时)

冲浪者关键词表：这是由选择该关键词之后单独用户找到的有用的关键词表组成的数据集。这是通过跟踪该用户决定使用哪个关键词找到的。这等同于冲浪者跟踪。

输入数据集	输出数据集
<ul style="list-style-type: none"> · 关键词(临时) · 有关从关键词建议者使用什么关键词的数据 	<ul style="list-style-type: none"> · 该单独用户找到的有用的(与关键词相关联的)其它关键词的排列表(半永久)

关键词建议者：这是由用户已找到的有用的，从连续冲浪者关键词表编译的其它关键词的永久排列的数据集，并链接到每个关键词(这等同于该用户的选择的命中表)。

输入数据集	输出数据集
· 冲浪者关键词表(临时和永久) · 现有用户的选择的命中表(永久)	· 新用户的选择的关键词表(永久)

用户基本搜索算法

上面提供的讨论提供了更全面地描述本发明所需的语言。如图3A和3B所示，该图提供了根据本发明的搜索引擎能力的概要，其中在形成提供给终端用户的搜索结果的过程中选择网页。在步骤112，用户输入多达4组数据：关键词52，简介类型54，搜索类型58和用户ID56。IP地址62和日期时间60不由用户输入，但用户使用该搜索引擎时可读出。在步骤114和116中并行使用该数据以产生网页列表。下文详细讨论的步骤114是从根据本发明产生的新颖的新搜索引擎数据集选择网页的过程。如果需要，这可与步骤116并行运行，步骤116获得从其它现有搜索引擎选择的网页。此后，组合从步骤114和116选择的网页并在步骤118标记。下面更详细描述标记网页列表的处理可生成图3中冲浪者跟踪数据所示的数据集并在搜索引擎用户从步骤120中的列表选择网页时发回到搜索引擎。选择标记号的网页的处理生成后面的数据系列，用该数据系列更新搜索引擎数据集：关键词124，URL126，用户ID130，日期时间132，简要网页说明134。

虽然最好是使用冲浪者跟踪数据中的所有这些不同的数据类型，使用该数据的不同组合完全在本发明的要求范围内。当新网点加到搜索引擎10的数据集114时，描述134通常仅包括在本发明的优选实施例中，所使用的描述将是出现在原始网页列表上的描述。如

下面进一步说明的，日期时间数据132可仅表示一个网点被选择，而不是记录用户在特定网点的时间周期。在从网页列表选择网页时，该过程是在步骤122直接占用对应的URL的用户看不见的。下面更详细地描述步骤114，118和120的实施细节。

在初始选择之后，用户可选择访问另一个网页URL搜索结果。根据该网点的关联，用户可以费时读取，下载，探索进一步的网页，嵌入的链接等时间，或如果对该网点的出现不关联/不感兴趣，在短周期后用户可以直接返回该搜索结果。记录两个选择之间的时间差作为与来自网页搜索表的后续选择的两个日期/时间数据132之间的差(在该实施例中，如果在访问该网页后做出另一个选择，人们可仅测量在一个网页花费的时间，然后将其提供给允许计算时间差的另一个冲浪者跟踪132)。有关该网页流行性的冲浪者跟踪数据被用来排列后面的搜索，如下文进一步描述的。

因此，根据本发明，推论和分析将其用来建立与该搜索主题的不同结果的关联是人类用户的能力。本发明采用所有人类用户的累积处理和推论，以便提供比上面可能的方法类型获得所需信息源有效得多的手段。

如上所述，通过在每个关键词搜索后记录用户去了哪个网页来捕捉人脑的力量。根据本发明，通过在对用户搜索产生网页列表中把自动发送信息的隐藏链接发回到搜索引擎(或辅助服务器)实现对冲浪者跟踪数据的集中。虽然用户仅看到显示了其要求的链接，隐藏的链接通知可用Java小应用程序执行处理的转移的搜索引擎。因此，当因特网用户选择一个网页时，它将用户引向该地址，但是还向搜索引擎10发出指出已选择了什么的冲浪者跟踪数据。当用户返回到网页列表和选择另一个网页列表时，则执行生成另一个冲浪者跟踪的另一个Java小应用程序。来自两个后续选择的该冲浪者跟踪中的数据时间数据之间的差别捕捉该用户已在前一个网站的时间周期。用户不了解该数据被发出的情况发生。

在另一个实施例中，不是使用多个Java小应用程序集中冲浪者

跟踪数据的完整列表，没有描述数据134，日期时间数据132表示一个用户访问了一个特定的网站。在一个具体的实施例中，用户必须访问特定的网站比预定的时间周期多，例如1分钟或15分钟，取决于什么是已查看该网点以便对访问的网点计数和向搜索引擎10发回任何冲浪者跟踪数据的适当时间，同样如下文所述。在该实施例中，每个小应用程序包含在搜索引擎更新数据库所需的所有信息。另一个实施例在用户导航到要求的网站之前集中冲浪者跟踪数据。获得该冲浪者跟踪数据的其它方式也是可能的并且在本发明所要求的范围之内。

因此，根据本发明的搜索结果页是与常规的搜索引擎的结果页不同形成的。其差别在于作用而不是内容。在视觉上，对用户来说，该页看起来与来自其它搜索引擎的标准搜索结果相同。

一个实例说明这一点：在常规搜索中，对关键词"Weather"搜索的结果页可以读出：1.www.weather.com 今天的天气预报。预期今天各地晴好。

与"www.weather.com"标记相关联的HTTP链接是"http://www.weather.com"。这表明：如果用户选择该链接，直接将他们导航到该网页。

相反，根据本发明，使用关键词"Weather"进行的搜索的标记的结果网页可以读出：1.www.weather.com 今天的天气预报。预计今天各地晴好。

与"www.weather.com"标记相关联的HTTP链接是："link.asp?n=1."。因此，如果用户选择该链接，在用户看不见的处理中，根据本发明首先将用户引导到与搜索引擎10的网络服务器对应的网点上的link.asp网页，并传送值为1的参数n。

服务器侧的代码(在万维网服务器上运行的应用代码)使用该参数识别URL和用户的选择网点的描述。然后将该信息与其它冲浪者跟踪数据一起存储在数据库表中。然后服务器侧代码对用户需要的URL执行改向操作。用户则看到其所需的网页出现。

搜索结果的来源与该活动无关。用户的目的地网页与该活动无关。该处理是将用户，关键词和目的地记录到数据库中的一种。该跟踪方法可仅记录关键词搜索后最初访问的网页。如果用户连续返回到该搜索结果表，则可记录后面的网页访问。

可对作为用户选择网页的结果发回到搜索引擎10的数据集114的冲浪者跟踪数据加密，以防止欺诈的用户把假数据发回到搜索引擎。

用户可从最初的URL选择连接到(如果他们不返回搜索结果页)的另一种跟踪方法是运行所选择的网页作为位于搜索引擎网站的'帧'的部分。这样允许在输入关键词后记录所访问的网页的完整记录。然而，这样会另外增加系统的复杂程度，可能会降低系统响应时间。

如前面提到的，由于其提供了搜索者(冲浪者)如何使用搜索引擎的跟踪或记录，可收集的冲浪者跟踪数据包括关键词124，URL126，用户ID128，IP地址130，日期时间132，简要网页说明134，并这样来识别。该数据用来改善在前面的搜索者的喜好上建立的未来的搜索。因此，冲浪者跟踪是单独用户或万维网'冲浪者'针对特定的关键词组从初始搜索结果优选的一种手段。如何生成确定网页列表的数据集合。

图4说明了根据本发明用于不同网页搜索的数据集。用于确定该网页列表的数据集包括关键词表164，简介ID表166，安全表168，累积的冲浪者跟踪表170，关键词URL链接表172，个人链接表174，和网页(URL)表188。

下面更详细地描述上述数据集的结构。后面的描述给出以电子表格方式排列的数据，每个单元有多个值并多个空白单元。以这种方式进行说明便于解释本发明，但不是一种有效的存储和检索方法。关系数据库模型用于根据本发明实现数据存储，以致有多个所涉及的字段或表来存储数据对于本领域的那些技术人员来说是很明显的，每个字段仅存储一个值。

关键词表(164)

在下面所示的表1中更详细地给出图4的关键词数据表164的内容，并且是关键词表，包括短语，和已向他们请求的次数。如果该表变得较大以致不能管理，可从该列表删除预定时间周期之后不再使用的关键词。然而，如果可能，希望保留输入的大部分或所有关键词短语。

关键词	该关键词被请求的累积次数(W)	每个关键词的唯一编号
关键词1	W1, W2, W3等	
关键词2		
关键词3		
关键词4		
关键词5		
关键词6		
关键词7		

表1 信息请求列表和其被请求的次数

可根据所选择的(W1, W2, W3, ...), 例如W1=总搜索, W2=男性简介, W3=女性简介, W4=USA简介等不同的"用户简介"来分离一个关键词被请求的累积次数。应指出, 由于用户可落入多于一个简介类别, 例如来自USA(W3)的男性(W2), W的总和比一个网点已被访问的总次数大。这不仅将变成使用该关键词的用户搜索者的数量, 而且是用户(根据所选择的简介类型)搜索该关键词的类型表。如下文所述, 虽然他们涉及使用关键词建议器, 只要拼写不同, 表明相同事物的关键词在不同语言中是不同的关键词。

网页表(188)

下面所示的表2中更详细地给出了图4的网页表188的内容, 并包含因特网网页列表。每个网页具有URL地址, 相关联的2-3行说明, 对每个URL(也可以是任何字符, 符号代码或表达式)唯一的网页数量, 和该URL已被访问的累积次数。URL地址具有向其分配的唯一

编号(也可以是任何字符, 符号代码或表达式), 而不是在后面的数据表中存储的完全URL串。

地址	2-3行说明	每个 URL 地址的唯一编号	URL(网页)被访问的频率
URL地址1			
URL地址2			
URL地址3			
URL地址4			
URL地址5			
URL地址6			
URL地址7...			

表2是信息提供者列表和网页说明

关键词URL链接表(172)

下面所示的表3中详细给出了图4的关键词URL链接表172的内容。由于该表包含信息提供(URL地址或网页)和信息请求(关键词)之间有关链接的信息, 它对于本发明来说特别重要。

该数据将来记录在描述关键词与由下面的三个参数定义的具体值之间的关系的数据集。

-对与每个关键词对应的每个URL地址的有效访问(命中)的累积数量(在此称之为X或加权系数X)。这是每个关键词的URL的流行性的量度并从冲浪者跟踪确定。

-在更早的预定时刻测量的有效访问的以前的累积数量(在此称之为Y或加权系数Y)

-与每个所述网页的生成或输入时刻有关的日期时间系数(在此称之为Z或加权系数Z)。Z是网页开发者向搜索引擎提交网页的日期时间。

并不是关键词和URL地址的所有组合都具有数据X, Y和Z。

	关键词	关键词	关键词	关键词	关键词
URL地址1	X,Y,Z				
URL地址2					X,Y,Z
URL地址3			X,Y,Z		
URL地址4	X,Y,Z				
URL地址5		X,Y,Z		X,Y,Z	
URL地址6					
URL地址7					

表3是信息提供者(网页)与信息请求(关键词)之间的链接
具有该关键词URL链接表的简介类型

网页的流行对于不同的人群是不同的。包括多种简介类型将产生表3中的多个X, Y和Z的值, 例如人们可具有由X1 X2 Y1 Y2等表示的全球和新西兰的流行速度。

	关键词"体育"
与橄榄球有关的URL地址	X1=520, X2=52
与篮球有关的URL地址	X1=4000 X2=20

在该例中, 橄榄球和篮球URL地址的全球流行性(使用一般的简介类型)分别是520和4000, 对于新西兰的简介类型分别是52和20。

当使用一般的简介类型设定时(根据X1排列), 篮球网点排列在顶部。当选择新西兰设定时(根据X2排列), 橄榄球网点将是最高的。这反映了新西兰人的喜好。这是一种存储不同人群的喜好的非常简单的方法。

人们会期望基于新西兰的橄榄球网点在新西兰表上比海外网点评定高, 但没有必须是这种情况的原因。西班牙在世界上可能具有最好的橄榄球网点。该系统仅对用户感觉的信息质量评估网页, 网点的物理位置并不重要。

表示不同国家, 职业, 性别, 年龄等的X值有非常大的范围, 能够非常简单地捕捉不同人群的流行性。用户可根据其个人的兴趣/特征选择组合任何X值。

作为例子, 如果说,

- X1表示男性
- X2表示女性
- X3表示新西兰人
- X4表示美国
- X5表示工程师
- X6表示律师

一个"男性"和"新西兰人"将使用X3和X1两个搜索引擎增量。该便利条件增加了系统的数据需求, 但它对不同的用户将极大地改善搜索结果。由于用户可加入一种以上的人群, 因为用户可属于一种以上的简介类型, 需要将网页的总流行率作为分开数量存储。所有单独流行率之和将大于总流行率。

为针对用户简化该系统, 存在一种缺省简介类型(X的选择), 该缺省简介类型具有一个使用其它简介类型进行特定搜索的选项。例如, 用户可具有新西兰男性的缺省简介类型, 但如果需要一种技术, 可选择一个反映全世界工程师的累加的搜索知识的'全球工程师'简介类型。

人格化的程度取决于搜索的频率。例如, 诸如"新闻"之类的常用关键词将具有高人格化程度(大范围的X值), 如"英语邮票"之类较不太普遍的关键词有很少或没有人格化(仅有全球X值)。人格化的程度是使用该关键词(从表1找到的)的频率的函数。

累积冲浪者跟踪表(170)

下面所示的表4更详细地给出图4的累积冲浪者跟踪表170的内容。用冲浪者跟踪数据更新与网页和表3中的关键词(也称为关键词URL链接表172)之间的链接有关的信息。累积的冲浪者跟踪是从所有独立的冲浪者跟踪组合的信息, 用它来确定"命中"多少(有效访

问), 每个网页针对每个关键词。

从每个独立的冲浪者跟踪收集的信息是前面描述的一串输入, 并在下面以表格的形式给出

IP 号码	用户 ID	关键词	URL(网页)	日期时间

表4中每一行是一个冲浪者跟踪, 组合的行是累积的冲浪者跟踪
下面进一步处理冲浪者跟踪数据以更新表3的方式

简介ID表(166)

下面所示的表5中更详细地给出图4的简介ID表166的内容。该表包括唯一标识, 密码, 联系电子邮件和他们通常用来进行搜索的缺省简介类型。

用户标识	密码	电子邮件	缺省简介	其它信息
Joe Bloggs	dogs	jbloggs@AOL	US, 男性	

表5是用户识别表

存储用户缺省简介类型作为用户的个人喜好简介部分, 通过向系统输入某些个人识别的形式来访问该简介。当登录到数据搜索引擎可提供该信息, 或如同本领域中已知的项目, 搜索引擎可在计算机上留下"cookie"以识别用户, (应有一个任选的电子邮件地址和与登录过程相关联的密码(或类似内容)。IP地址本身不是充分的识别手段, 由于它对各个用户来说不一定是唯一的。

其他信息可以包括用户定义的喜好如何对搜索结果进行组合以及特定用户感兴趣的关键词。该信息可用于主动地定制访问网页的搜索结果和建议。

个人链接表(174)

下面所示的表6中更详细地给出图4的个人链接表174的内容。表

6与表3的结构相同，并可用来记录与包括访问次数和关键词的每个URL相关的用户个人喜好。然而，在该表6中，Z不是网页开发者提交该网页的日期，而是用户访问该网页的日期时间。这样允许用户通过定义他们访问的网页的最后时间可精选搜索。

	关键词	关键词	关键词	关键词	关键词
URL地址1	x,y,z				
URL地址2					x,y,z
URL地址3			x,y,z		
URL地址4	x,y,z				
URL地址5		x,y,z		x,y,z	
URL地址6					
URL地址7					

表6是信息提供者(网页)和一个单独用户的信息请求(关键词)之间的链接

表6中的数据可由生成它的个人访问，并可使用用户ID访问，它最好与用户的电子邮件的改变或IP地址的改变无关，并因此能在该变化期间保持其过去的个人喜好。

表6的数据集可存储在搜索引擎网点或个人计算机上。存储在本地PC上将需要在用户计算机上安装附加软件。在搜索引擎存储该信息有许多优点，包括用户更经常喜欢去的地方，一旦他们有了一个基本的书签表就不喜欢改变搜索引擎。

安全表(168)

下面所示的表7中更详细地示出图4的安全表168的内容。为确保用户不一次又一次地提交相同的关键词以增加其普及，使用下面的安全数据表。每一项是单一的信息段，即是或否。可针对关键词和IP地址之间的链接或关键词和用户ID之间的链接生成该表。

	关键词1	关键词2	关键词3	关键词4
IP地址1	1			
IP地址2				1
IP地址3				
IP地址4				1
IP地址5		1		

表7是确保一个计算机用户不提交人为地提高一个网页的流行率的关键词

下面描述本发明使用的处理以便占据前面提到的图4表的每一项。

传播关键词表164

每当用户向搜索引擎输入关键词52时传播该表。把提交的关键词与表1中的关键词列表(关键词表164)比较,如果其仍未出现则将其加入。如果其已出现,将累积数量加1。如果用户具有简介类型,则增加每个简介类型的关键词的累积数量(W1, W2, W3等)。

传播网页数据表(URL表)188

以许多方式传播该表,包括:

- 用户选择表2(URL表188)中仍没有的URL地址126。将URL地址126和说明134直接放入网页数据表188。向新URL分配唯一的识别编号。

- 在如图4所示的步骤176,网页开发者可提交同样直接进入网页数据表188的URL187和说明68。

- WebCrawler也可增加URL地址和说明(该说明是该网页的前几行或在HTML编码的"标题"中)。这不是该系统的主要单元,但它可能是获得URL和说明的方法。利用该搜索系统,更有可能使用WebCrawler检验信息而不是找到新信息。

传播累积的冲浪者跟踪表170

每当由用户选择"标记的"网页时,传播累积的冲浪者跟踪表170,也是指上面的表4。发送一个冲浪者跟踪信息分组,以致每当用户从网页列表选择另一个网页时将该冲浪者跟踪数据加到该表。

传播关键词URL链接表172

用来自累积冲浪者跟踪170的数据更新网页的流行率,记录到表3(X, Y)中,也称为关键词URL链接表172。用来自累积冲浪者跟踪(170)的数据更新表3以获得X和Y的新值的频率是可改变的变量,该变化在比每小时一次短到比每月一次长的范围内。应指出,在不同的时间间隔更新不同关键词。

处理中的一个中间步骤,累积的冲浪者跟踪形成累积的冲浪者命中表。以后用其修改表3中X和Y的值。

如上面提到的,记录关键词和URL之间的链接("有用的访问"或"命中")的最简单的方法是对每个关键词,作为"命中"的冲浪者跟踪中削减的URL计数。一种更有意义和先进的方法是:如果该用户满足特定标准,仅对作为有效的位置选择计数。该标准可能是用户在一个位置超过特定时间。如果不满足该标准,该选择不增加表3中累积的X的值。

也可以根据在网页度过的时间增加X的值。度过的时间越长,X的值增加得越多。X不必是整数。

由于网站能力在登录时间,下载时间,带宽,和响应时间方面的变化,可适当地改变用于表示有效'命中'的预定时间。可采用专业WebCrawler以使该数据独立有效。

在关键词搜索对其网页的命中计数(增加X的值)后选择内容提供者的标志。这样能使其网页在与该关键词相关联的流行表中上升。它作为网页开发者使一个关键词值得被查看的机理。他们可以不付费上升到流行表上,这仅在如果人们访问其站点并在那里度过时间和记录对该流行表的有效命中时出现。内容命中值(例如如果是1或0.5或7)可根据人们想使该内容对流行排列有多少影响的重要性而改变。

可以以许多种方式处理累积的冲浪者跟踪信息以占据表8(下文中)。根据IP地址或用户ID将累积的冲浪者跟踪分组产生了针对各个用户的搜索形式。这是一种关键词和URL和时间的列表。允许计算每个用户在每个网页度过的时间(不能计算在一个搜索任务的最后的网页度过的时间,因为在他们到该网页后没有时间记录)。

如果每次访问之间的时间比特定的时间周期长,向关键词URL的累积冲浪者命中(α)表加1,(这是最简单的方法,例如,相关性与在该站点度过的时间成比例的方法也在本发明的范围内)。

	关键词	关键词	关键词	关键词
URL地址1				
URL地址2	α		α	
URL地址3		α		α
URL地址4		α		
URL地址5				
URL地址6			α	
URL地址7				α

表8是从累积冲浪者跟踪生成的累积冲浪者命中表

以下面的方式用累积冲浪者命中来更新表3中的X值

$$X_{\text{(new)}} = (X_{\text{(old)}} \cdot \text{HF}) + \alpha$$

HF是一个数值在0和1之间的历史系数。该历史系数对每个关键词来说不必相同并可根据使用该关键词的比例来改变。

在预定时间周期之后,使用对表8收集的数据重新计算表3中X的值。更新表3的频率将影响所选择的历史系数(HF)的值。把现有的X乘以"历史系数"的原因是使感觉的流行不是无限地持续。历史系数减小了附加到过去流行率的加权。作为例子说明,关键词"体育新闻"可具有按下面排列的现有流行率(根据每个网页的命中数量X)

- | | |
|------------|---------|
| 1 冬季奥运会网页 | X=19000 |
| 2 足球比赛结果网页 | X=18000 |
| 3 棒球比赛结果网页 | X=15000 |

4 高尔夫新闻网页 $X=15000$

一个星期的累积冲浪者命中表可以是:

1 冬季奥运会网页 $\alpha=500$

2 足球比赛结果网页 $\alpha=1800$

3 棒球比赛结果网页 $\alpha=1500$

4 高尔夫新闻网页 $\alpha=4600$

命中数量的改变反映了冬季奥运会已结束和名人高尔夫锦标赛已开始的原因。如果人们具有一个0.9的"历史系数", 新流行率(X)将是:

1 高尔夫新闻网页 $18100(0.9 \times 15000 + 4600)$

2 足球比赛结果网页 $18000(0.9 \times 18000 + 1800)$

3 冬季奥运会网页 $17600(0.9 \times 19000 + 500)$

4 棒球比赛结果网页 $15000(0.9 \times 15000 + 1500)$

因此, 较流行的网页上升, 不太流行的网页下降, 反映了对不同题目和事件的兴趣在一段时间内的波动。

因此, 该数据库利用人的意志来提供人们在因特网上找到有用内容的有力表示。用户本身通过该搜索来替换在其它情况下滤除所需的基本计算要求。

表3中Y的值是旧的X值, 并以认为是适当的间隔更新Y的值, 该间隔可以是分钟, 小时, 天, 星期或更长。如前面提到的, 对不同的关键词来说, 更新间隔不必相同。用其计算网页流行性改变的速度并可用作选择标准。

网页/URL链接表中的不同简介类型

累积的冲浪者跟踪包括有关用户的信息, 因此可用不同简介类型 α 的下标值计算表8。这些 $\alpha_1, \alpha_2, \alpha_3$ 的值对应于X的下标值的简介类型。这样允许记录不同人群的流行率。

输入到网页/URL链接表172的新网页数据

对于网页开发者来说, 用搜索引擎记录新网页的最简单的方法是每当他们生成或更新网页时提交如图4中的作用176所示的信息,

该信息包括URL66, 关键词70, 站点说明68, 目标读者72, 和日期时间74。

该信息直接更新表2(图4的URL表188)和3(图4的关键词URL链接表172)。对于每个关键词(70), 在表2中输入URL66和说明68, 在表3中插入提交该页的日期时间(74)。用户被允许他们可用其提交其网页的一组关键词70。下面给出仅有Z值的表3看起来象什么的例子(格式: 日/月/年)

	关键词	关键词	关键词	关键词	关键词	关键词
URL地址	27/02/98		27/02/98			
URL地址	28/02/98	28/02/98		28/02/98		
URL地址						
URL地址			18/02/98		18/02/98	18/02/98
URL地址						
URL地址				28/02/98		
URL地址		29/02/98				

表9是从万维网开发者提交的内容生成的数据表

如果表3中没有URL和关键词组合的日期, 则自动插入新日期。如果表中已存在日期, 则对这些日期进行比较, 如果日期靠得太近, 即预定周期, 则保留旧日期并忽略新日期。这样阻止了人们不断地重新提交, 来通过重新提交其网页达到在新网页表顶部的目的。如果表3中的URL具有比预定周期更近的Z值的关键词, 则不允许该提交。这样阻止了网页开发者利用不同的关键词组重新提交其网页。

当用户提交URL时, 他们可按表3将其目标定在特定的用户类型(不同的简介类型Z1, Z2, Z3等)。例如, 当新西兰人搜索该关键词时, 专门以新西兰人(例如Z1)为目标的URL提交将出现在关键词新列表的顶部。它保留在顶部直到某人另外提交以新西兰人为目标的该关键词。以其它读者为目标的URL不作为针对新西兰人的新站点出现, 或作为替换, 他们不在专门以新西兰人为目标的新表中的高度为特征。

网页开发者不必输入有关新网页的数据。每当有关网页的信息已明显改变时，通过向搜索引擎自动提交数据的万维网文献模板可自动进行。提醒网页开发者改变成任何适当的关键词。

虽然现有的WebCrawler有许多缺陷，另一个实施例需要发出专业浏览器，以找到万维网站点地址和关键词。仅当万维网设计者用识别的关键词仔细构成其网页时他才是有效的。可推测喜欢进行这项工作的任何万维站点的设计者/所有者愿意向搜索引擎提交任何更新，以便从有关搜索结果的瞬时表得到好处。

该原理的延伸是自动检测一个万维网地址是否拥有数据库中的关键词信息，然而自动发出邀请，以便提供能使其很容易地找到其网页的信息。由每个网页提交的关键词的理想数量是最好低于50并且最好是在约5和20的范围内。这样也有利于迫使网站设计者找到最合适的关键词描述其站点，并且也能使他们选择他们希望作为目标的读者。

网页提交过程也可包括限制人们使用该系统的欺诈能力的网页开发者识别处理。它可包括防止多网页提交的惩罚。

传播简介ID表166

从来自用户的直接输入传播图4的ID表166。当用户搜索时，可从分层的下拉菜单选择其简介类型54，该菜单可包括，例如：

- 性别(男性或女性)
- 职业(专业，学生等)
- 年龄级别等

用户从该选项中选择不同的简介类型，如果他们想保存该类型作为其缺省简介类型，则对他们进行提醒。然后将其记录在表5中(表166的简介ID)。用户也可从专门的个人化选项页，而不是搜索页的下拉菜单上选择个人化选项。

传播个人链接表174

根据用户ID126的存储用累积的冲浪者跟踪识别各个用户的搜索模式。以与用累积的冲浪者跟踪170更新表3(关键词URL链接表

172)相同的方式更新个人链接表174。该表按自动书签的形式存储用户过去的喜好。

传播安全表168

每当用户将关键词52输入到搜索引擎时，通过在关键词52和IP地址62之间进行链接(或在关键词52和用户ID56之间进行链接)来更新安全表168(表7)。定期更新表7中的数据，其目的是阻止有系统的重复搜索影响流行表(表3中X的值)，而不是阻止偶然进行重复的关键词搜索的各个用户影响该流行表。

确定网页的表

图5示出前面描述的各种数据集合，提供了引起所建议的网页表中的各种输入和作用，并在下面更详细地描述。如图5所示，输入到搜索引擎的用户数据可包括：关键词52，用户ID56，搜索类型58，IP地址62，简介类型54。下面描述为生成冲浪者跟踪的目的如何使用该数据确定网页表250以及决定标记哪个网页表(图3的步骤118)。

表3中对应于图5中的关键词URL链接表172的数值(X，Y和Z)包含给出下面的搜索类型58所需的所有信息：

- 根据数值X的该关键词最流行的URL的流行表搜索排列命中表
- 根据日期/时间(Z)针对该关键词的最新URL的热度减弱新闻搜索排列的命中表
- 根据X和Y之间的差上升最快的URL的高跳跃搜索排列的命中表
- 具有数值X，Y或Z中任何一个的URL的随机取样的随机搜索命中表
- 数据生成搜索，这是根据日期时间Z和用户感兴趣的指定日期(不只是最新的)的命中表

个人链接表174还允许列出过去的喜好作为搜索结果

· 前面喜好的搜索是根据前面各体(来自表6的X)的流行率排列的命中表。该搜索仅以各个用户以前的搜索为基础。这样允许用户非常迅速地找到他们以前访问的站点。

还可使用许多其它搜索选项

- 常规搜索是来自普通搜索引擎的搜索结果表(图3中的116)
- 仅搜索其它内容。这是其它内容的表, 例如与该关键词相关联的广告。

可以用许多不同方式组合这些搜索结果

- 作为上述任何搜索命中表的集合的集中搜索排列命中表(这是缺省的搜索结果组)
- 可以是用户定义的任何上面的表的组合的定制搜索排列命中表。

图5还说明在决定246中使用关键词表164和安全表168以发出标记的网页。该决定基于关键词的使用频率, 安全表中的数据和用户标识的出现。图16中完全描述了发出标记的网页的决定的细节。

如何实施不同类型的搜索表

下面提供有关如何实施这些搜索类型中的每一种的更多详情以及每种类型的某些优点和缺点。该系统依赖于用户头脑的力量, 此时确定他们想进行何种搜索, 这取决于他们想找到什么。很容易说明该搜索方法, 因此用户应直观地了解到使用哪一个。

流行搜索

图6说明了确定与步骤272中输入的关键词270相关联的流行网页表的过程。如果选择该搜索并输入一个关键词, 步骤274随后, 并针对输入的关键词270根据从表3(图5的172)取出的X值产生网页表。用来自表3的唯一网页(URL)编号识别这些网页。此后, 在步骤276, 把从步骤274找到的网页编号表与来自表2(图5的188)的URL地址和网页说明组合。然后, 在步骤278, 如上所述依据图5中步骤246的结果来标记得到的网页表, 并将他们发送到用户以进行其选择。热度减弱新闻搜索。

图7说明了确定与步骤290中输入的关键词相关联的新网页表的过程。如果选择该搜索并输入一个关键词, 步骤294随后, 并针对在步骤290输入的关键词根据从表3(图5的关键词URL链接表172)取出

的Z值产生网页表。用来自表3的唯一网页(URL)编号识别这些网页。此后,在步骤296,把从步骤294找到的网页编号表与来自表2(图5的URL表188)的URL地址和网页说明组合。然后,在步骤298,如上所述依据图5中步骤246的结果来标记得到的网页表,并将他们发送到用户以进行其选择。

在提交每个网页时,用户还能够准确地看到,因此因特网用户可了解其流通。该特性的一个间接结果是鼓励万维网设计者更新其站点。给出新的和更新的站点突出提供了在流行命中表上建立的合适的手段,并鼓励使用适当的关键词和奖励用户找到的有用的网页保持上升。

高跳跃搜索

图8说明了与步骤320输入的关键词相关联的高跳跃网页搜索。这是流行率提高最快的网页表。如果选择该搜索并输入一个关键词,步骤324随后,并针对在步骤320输入的关键词根据从表3(图5的172)取出的X和Y值之间的关系产生网页表。用来自表3的唯一网页(URL)编号识别这些网页。此后,在步骤326,把从步骤324找到的网页编号表与来自表2(图5的188)的URL地址和网页说明组合。然后,在步骤328,依据图5中步骤246的结果来标记得到的网页表,并将他们发送到用户以进行其选择。

通过把来自表3的旧流行排列(Y)与新流行排列(X)比较来计算高跳跃。由此计算命中百分比的增加。另一种方法是按他们与上一次相比上升的位置数排列流行的变化率。

下面给出计算该实施例的流行变化率的公式:

$$((X-Y)/Y).(X/(X_m \beta))$$

其中 X_m 是针对对应的关键词的X的最大值, β 是可改变的附加变量,以便在流行表的顶部和底部改变相对明显的变化。

倍增X的最大值的原因是确保在较低流行等级的较小变化不会淹没在该表中上升较高的更明显的变化中。例如,以前仅有一次选择记录而此后的第二天吸引5次命中的网站会呈现出增加500%的百

分比,而另一个网页可能已经历了从520次命中增加到4000次命中(一个非常明显的增加),虽然这在另一方面表现出较低的增加百分比。

随机搜索

这是想看到除陈旧的跟踪外的用户根据具有与输入的关键词相关联的X, Y, 和Z的任何值的网页的随机选择对不流行的网页的随机选择。因此,用户在如图9所示的步骤352输入关键词之后,对图5所示的关键词URL链接表172进行参考,如步骤354所示,使用随机数发生器确定自动生成的随机网页表。在该随机选择中仅选择具有与关键词相关联的X, Y或Z值的网页,这表明在过去的某个阶段使用了该网页或网页开发者认为该网页与该关键词有某些联系。此后,在步骤356,把从步骤354找到的网页数表与来自表2(图5的188)的地址和网页说明组合。在步骤358,如上所述依据图5中步骤246的结果来标记得到的网页表,并将他们发送到用户以进行其选择。

常规搜索

这是常规搜索引擎的正常搜索方法,是指图3中的其它搜索引擎,如前面指出的,在用户选项中,该搜索方法可包括或不包括根据本发明的搜索。

仅搜索内容

这是一个内容表,例如与关键词相关联的广告,用户不能对其进行控制。如下文进一步说明的,根据本发明的优选实施例,付费最多的内容在该表顶部。当然,也可执行识别付费内容提供者的等级的其它系统。

以前喜欢的搜索

图10表示以前喜欢的搜索,它仅基于各个用户以前的搜索。这样允许用户非常迅速地找到他们以前访问过的站点并因此自动做出书签。应该指出,由于最好是使用密码登录根据本发明的搜索引擎系统,用户可从任何计算机访问其个人喜好。

因此,当用户在如图10所示的步骤372输入关键词时,其后是步骤374,在该步骤期间,从图5所示的个人链接表174(根据以前的使

用)确定该关键词喜好什么站点。由于用户具有可用来登录到该系统上的密码,因此,用户能够从任何计算机访问其个人喜好形式。

由于该搜索能力,因而不需要手动地做网页书签。如果用户忘了在好站点,例如'marbles'上做书签,他们通过重新打入将他们引到该站点的关键词可很容易地找到该站点。如果用户的喜好改变,他们将反映在个人链接表174中。

个人优选搜索的另一个实施例包括使用或不使用关键词规定最后一个访问该网页的日期。然后根据图5的个人链接表174中的Z来排列这些网页。例如,如果用户在上一年的一年中看到一一个站点,该用户可通过日期精选该搜索,从而使其容易找到以前有用的网页,即使他们不能记住相关的关键词。

该自动书签特性也可作为监视由特定计算机承担的因特网使用类型的设备,并因此可向访问不希望的站点,例如成人网页的儿童/雇员的父母/雇主例如提供警告。在优选实施例中,对于他们自己不喜欢使用计算机的父母/雇主,由计算机向父母/雇主自动提供儿童/雇员选择的关键词和访问的网页的使用通知。从图5的每个用户表174直接找到该信息。这需要也包括父母/雇主信息的用户识别码。

集中搜索

如图11所示的集中搜索是根据本发明的缺省搜索,并在用户不主动选择其它搜索选项时使用。

在步骤402输入关键词时,用该关键词从与该关键词相关联的网页选择组合中选择。如图所示,例如,在步骤404,用传统,流行,高跳跃,新和过去搜索结果的同等加权的组合获得网页数的表。此后,在步骤406,把从步骤404找到的网页数量表与来自表2(图5的188)的URL地址和网页说明组合。然后,在步骤408,如上所述依据图5中步骤246的结果来标记得到的网页表,并将他们发送到用户以进行其选择。首先构成该系统,搜索引擎10的数据库不拥有有关流行,高跳跃和新网页命中表的任何信息,因此搜索结果最初是从传统命中表(正常搜索引擎)获得的,然后用标记号的网页生成已描述的数据

库集合。随着系统开发，与其它搜索类型中的每一种相关联的数据集合变得流行，使用其它搜索类型的搜索将变得更有用。

生成日期搜索

图12表示允许用户选择提交该网页的日期的生成日期搜索。该特性仅针对包含按图4中的日期时间提交74识别的生成日期的数据术的网页工作。在步骤432输入日期时间和/或关键词时，搜索引擎10执行步骤434，其中获得与这些变量相关联的网页数量表。此后，在步骤438，把从步骤404找到的网页数量表与来自表2(图5的188)的URL地址和网页说明组合。然后，在步骤438，如上所述依据图5中步骤246的结果来标记得到的网页表，并将他们发送到用户以进行其选择。

定制搜索

图13表示允许用户决定他们希望其缺省命中表如何出现的定制搜索。在步骤462，选择关键词和用户ID以便开始定制搜索。在步骤466开始定制搜索之前，该步骤与前面相对图11描述的集中搜索的步骤404相同，应用步骤464以定制命中表的用户缺省混合，例如，用户可希望其缺省搜索结果，以便仅包括流行的和新的网页，但不包括高飞网页。然后在步骤466执行该定制搜索以产生网页号码表。此后，在步骤468，把从步骤466找到的网页号码表与来自表2(图5的188)的URL地址和网页说明组合。然后，在步骤470，如上所述依据图5中步骤246的结果来标记得到的网页表，并将他们发送到用户以进行其选择。在一个优选实施例中，可通过从用户的行为学习来修改缺省搜索结果表的形成，以便根据用户自己的搜索模式生成一个改变的定制搜索。如果用户始终选择例如新网页或高飞网页，则改变其缺省搜索结果以反映其通常的搜索风格。

杂志搜索

根据本发明的杂志搜索能使用户通过遵循一系列菜单驱动的选择(或相似的分级结构)，而不是输入特定的关键词进行搜索。

现有的杂志风格搜索引擎需要编辑器以便设定信息结构，决定

其相关优点，和对发送到用户/观众的给定页的空间设定标准，例如价格。使用本发明的搜索系统，用户自身动态地决定什么值得观看和什么不值得观看。因此，虽然需要关于主题分级的编辑输入，将自动得出这些主题中每一个的最流行出现的网页。

使用不同人群的数据集合

可采用不同流行的命中表来提供反映不同文化，地区，职业，性别或年龄兴趣分结果。因此，如14所示，当用户在步骤490输入关键词和用户ID时，可使用用户的缺省简介反映同一"群"人希望看到的用户简介的网页类型。因此，在步骤494发生的搜索以从那些"群"加入的人的缺省简介获得的下标X，Y和Z值为基础，该缺省简介是按在步骤492获得的用户的个人简介识别的。因此，专为该用户识别的人群获得搜索结果，而不是全球的搜索结果。如前所述，从步骤496和498得到的网页结果表是专用于这群人的。

因此，对于具有简介类型的特定用户，选择新西兰作为地区系数，对田径运动和相关关键词，橄榄球资料的搜索可能以图形形式突出表示，而美国人的简介类型可产生偏向于例如棒球/美式足球的资料。该技术提供了根据与特定简介类型相关联的流行命中表的上下文内容鉴别相同文字之间的不同含义的能力。例如使用关键词'accommodation'的一般搜索会包括有关住房，租房等类似的结果，而如果用户在其简介类型中表示验光方面的兴趣，对术语'accommodation'的解释将完全不同。

自动推断这些站点的关系，而没有搜索引擎或用户对这些站点的任何主动评估。不需要复杂的算法来分析针对特性类型用户的网站的关系。而是由针对其简介类型选择那些特征的那些用户，即美国妇女喜欢攀岩来决定认为有关的站点类型。关系较大的站点自然吸引更多的命中，提高了其排位，从而增加了后面的用户也探索该站点的机会。在上面的例子中，在做出有效命中所需的时间周期不会访问为关键词'accommodation'列出的与验光，视力，透镜，视线无关的任何网站。因此，它将接受非常低的排位，并因此不太可能

被将来的用户访问。

在单一的任务期间用户可针对不同搜索选择不同简介类型并且不限于缺省简介类型。

在本发明的另一个实施例中，可包括特定人群以其搜索结果实际用于更新与该群人相关的数据库的目的的鉴别等级。例如，具有识别他们作为医生的用户ID的医生可进行与特定的医疗状况有关的搜索，可标记号其选择并在如以前所述的用于该群医生的数据库中使用。然而，虽然病人可能希望用同一群医生的用户ID识别其简介，其选择不象实际的医生的用户ID那样有效，因此，虽然他们能看到医生认为最适当的网页列表，由于其ID未识别他们是医生，不能用其选择来更新医生群的数据库。

有限搜索选项

本发明的另一个特性是图15所示的关键词限制器特性，当启动该关键词限制器特性时，防止如儿童之类的特定用户搜索不希望的关键词和网页。本发明人已认识到可能非常容易阻止儿童搜索例如色情内容，而不是尝试跟踪和防止访问因特网上具有色情内容的所有站点。这将用作现有的"net nanny"型设备的补充工具。因此，如图5所示，通过启动关键词限制器，把先存的不能访问的关键词表存储在一个表中，并在步骤522与预先输入的关键词比较，如步骤520所示。因此，不能搜索不可访问的关键词。因此，例如父母可选择他们不希望他们的小孩搜索的关键词类型，这对不同的父母组是不同的。该系统在步骤524滤除可能用于以后搜索的关键词。

确定哪些用户取样

图16表示确定应对哪些搜索结果取样的过程并用来形成图4的累积冲浪者跟踪表170，也称为表4。虽然是可能的，不必集中涉及每个单一搜索的数据，这可通过确定哪组结果发出带有"标记的"网页来控制。已相对于提供的例子中的医生的特定群的鉴别用户提到了对此的参考，

如图16所示，在步骤554输入关键词和其它数据后，存在三种确

定这些结果是否实际"标记了"的决定,如以前在图3的步骤118中描述的。

如步骤556所示,对于具有用户ID和已选择使用如前所述的图5的个人链接表174(表6)的用户,需要标记其所有结果,以便其过去的所有喜好记录在其个人链接表174中。根据本发明系统的搜索引擎可更新用户的个人喜好,但如果不满足特定的安全等级(见下文)不能更新表3。然而,如果个人链接表174存储在个人的计算机上而不是中央位置,由于本地存储该数据,不需要发出标记号的结果。

如步骤558所示,当提交关键词时,在用户发送一组标记的结果之前使用安全表168(表7)对仍未搜索的IP地址62进行检验。如果是这样,虽然该搜索对累积冲浪者跟踪(170(表4)不起作用,用户仍可承担该搜索。这样允许所有普通用户影响流行命中表,所有用户搜索他们喜欢的任何内容,但防止欺诈用户。例如spammer对流行命中表起作用。安全表168还可包括有关关键词52和用户ID56之间的链接的信息以检测重复搜索。

虽然用户能改变其计算机的IP地址,也可通过如登记和跟踪IP号码的使用的许多方法检测和防止。

排除错误搜索的其它方法包括:

- 仅对具有利用搜索引擎记录的用户ID554的用户生成冲浪者跟踪。
- 延长进行作为有用命中的访问计数所需的时间限制。
- 不对来自关键词的URL的单独访问计数(它不意味着测量经过的时间)。

如步骤560所示,每隔出现10次,100次,或甚至1000次跟踪流行关键词一次,可改变该选择频率以优化该系统。从如图5所示的关键词表(表1)确定关键词使用的频率。也可将发出标记的结果的频率链接到针对不同关键词改变流行率的比例。例如,关键词"IBM"可能会使IBM的主页在顶部,并且大部分用户会到这里,而关键词"latest fads"可能具有不断变化的需要更频繁取样的网页组。

为避免图5的关键词URL链接表172变得过大，一种方法是当关键词达到特定使用频率时仅将该关键词在表3中登记一次。这是通过对不太频繁使用的关键词(从表1找到)不发出标记的结果控制的。

访问主动建议的网页

图17表示本发明的另一个特性，并涉及使用数据以便主动建议网页。由于用户建立该请求，并且如果存在有关主题的任何新数据则通知该用户，它与搜索不同。为此，用户必须在步骤588主动规定他们对哪些关键词感兴趣和他们喜欢作为筛选器或媒介物的简介类型以及搜索类型(新，高跳跃，流行)。该信息存储在图5所示的用户简介ID166(表5)中。

因此，用户以各种间隔接收由同类想法的人群确定的所建议的网页表。例如，用户可通过下列内容选择要将通知的网页

关键词582	简介类型(媒介物588)	搜索类型586
橄榄球	新西兰，男性	高跳跃
腐牙治疗	牙医	新

这样，如果存在有关其他新西兰男性找到的有用的"橄榄球"的高跳跃网页(即他们花费大量时间查找该信息，表3中高变化率的X)，则通知该用户。同样，如果存在有关为查找的牙医提交的"腐牙治疗"的任何新信息，用户将对此进行识别(表3中的Z值)。计算机媒介不可能象所选择的同等人群那样擅长筛选信息。与其它"媒介型"软件相比，该系统的一个优点是不需要用户计算机上的任何软件。它全部包括到其它搜索引擎数据集的自然延伸。

当用户接下来访问该搜索引擎或选择通过电子邮件通知来向用户通知这些建议的网页时为该用户显示所建议的网站。这样，这些网页可引起用户的注意而不需要对这些关键词的任何主动搜索。

访问被动建议的网页

图18表示本发明的另一个特性，并涉及基于用户过去的搜索的

自动网页建议，并且不需要来自用户的主动输入。

如图所示，在步骤620，输入用户ID时，通过查看用户频繁使用图5的个人链接表174(表6)查看哪些关键词，简介类型和搜索类型来以各种间隔或时间(例如每次登录到该搜索引擎)被动地启动该系统。例如，用户可作为"新西兰男性"频繁查看橄榄球信息和作为"牙医"查看腐牙治疗。可从自动书签表，前面称为个人链接表174找到该信息。如果用户查看这些主题未达到特定的时间长度并存在新的或高飞信息来自，则自动向用户通知这些新信息来源。

在该实施例的一种改进中，可定期发出带有两个与用户的关键词相关的最新和最高升的站点的电子邮件。

确定建议的关键词表

对许多用户来说，用因特网搜索的一个问题是知道用哪个关键词进行搜索。虽然可以用无限数量的关键词实施本发明，用户选择太多关键词(包括短语)可能会出问题。

因此，如图19所示，本发明还提供数据集642，该数据集642提供在步骤640与特定简介类型一起输入的关键词的同义词。图19所示的系统被称为关键词建议器。在一个实施例中，它是通过把用户在步骤640输入的关键词与其它用户已尝试使用其他搜索方法的图5的关键词表164(表1)中现有的关键词和短语实现的，在步骤646中识别。然后在步骤660中标记每个关键词，用户在步骤662中选择的那些关键词被用来形成如图19所示的关键词冲浪者跟踪648，如同前面描述的网页冲浪者跟踪，关键词冲浪者跟踪包含用户输入的原始关键词52，选择的关键词652，和IP地址130，用户ID128以及日期时间132数据。

然后用来自累积关键词冲浪者跟踪648的数据增强关键词之间的链接。这样，系统了解到哪些关键词相互关联。该系统以与该系统了解哪些URL与关键词关联的相同方式了解哪些文字相互相关。由于每当使用关键词建议器时改进了其联系，所建议的关键词表在一段时间变得更加相关联。

生成确定所建议的关键词的数据集

如图20所示，关键词链接表696和累积关键词跟踪表698以及前面描述的安全表168被一起用来生成所建议的关键词的数据集。如下面的表10所示，关键词链接表696记录如何频繁地从所建议的关键词表选择每个关键词。然后可用此来排列相互有关的不同关键词的有用性。

	关键词1	关键词2	关键词3	关键词4	关键词5
关键词1	-	5			
关键词2	20	-	1134		
关键词3		356	-		
关键词4				-	
关键词5		20			-
关键词6			3		
关键词7					168

表10是关键词链接表

从表10可以看到，输入关键词2的人找到关键词3最有用，随后是关键词5，然后是关键词1。关键词可以有一个定向方面，例如，试过用户关键词2之后，关键词3被有效地找到1134次，然而，在用户试过关键词3后，关键词2仅被有效地找到356次。

用有关人们如何使用建议的关键词(关键词冲浪者跟踪648)的信息更新表10中的关键词之间有关链接的信息。累积的关键词冲浪者跟踪698是来自所有各个关键词冲浪跟踪648的组合信息，并用它来确定针对每个关键词多少次"命中"(有效访问)每个关键词。

从每个单独的冲浪者跟踪集中的信息是一系列变成累积的关键词冲浪者跟踪的输入，下面的表11中以表格形式示出。

IP号码	用户ID	关键词 (原始)	关键词 (建议)	日期时间

表11是关键词累积冲浪者跟踪

传播关键词链接表

图20还示出通过记录用户放入搜索引擎中的关键词如何能够开始表10关键词之间的链接。例如，如果某人使用关键词"NHL"，然后是"National Hockey League"进行搜索，通过将此作为一个命中记录来引起表10中两个关键词的关联。另外，这将捕捉用户的推理能力以定义两个关键词之间的链接。经常是后一个关键词与前一个关键词完全无关，但有时它是有关的。如果下一个用户从关键词选择器选择该关键词，它将以与对网页的重复选择来增强关键词和URL之间的链接的相同方式增强该关键词链接。

下面是输入象"Book"这样的简单关键词后可建议的关键词的例子

- 图书销售
- 图书综述
- 专业图书
- 二手图书
- 旧书
- 特刊图书

所有这些关键词(短语)信息来自搜索者(用户)和信息提供者(网页开发者)。最适合的关键词必然一直出现。

把用户使用的所有关键词输入到图20的关键词链接表696。因此，如果人们输入不常用的关键词"cassetes"代替"cassettes"，关键词建议器将建议用户尝试"cassettes"。因此，不需要在表3中为"cassetes"生成一组URL关键词链接。因而节省了数据空间并且还不

需要为关键词"cassetes"发送一组标记的结果。因此只向搜索引擎发回很少的数据。

所考虑的另一个实施例是运行象表4那样的关键词建议器并具有高跳跃关键词结合和新关键词结合，以便该系统可了解关键词结合一直如何改变。例如，关键词建议器跟踪可存储最近的关键词链接，并以与利用累积冲浪者跟踪修改表3的相同方式来利用历史系数修改主要关键词跟踪。

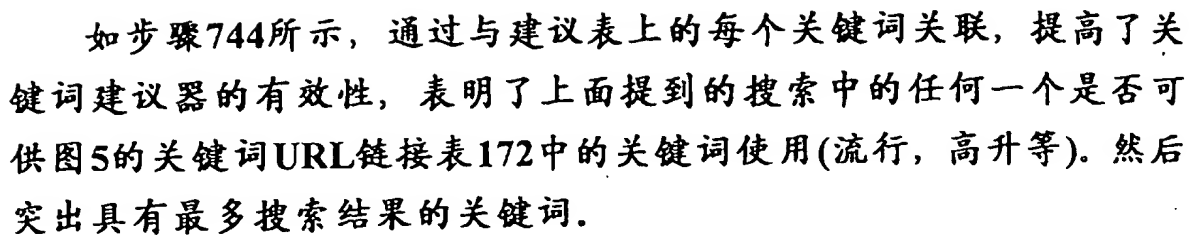
以与图5的累积网页冲浪者跟踪170相同的方式处理累积的关键词冲浪者跟踪698，以增强关键词链接表696(表10)中关键词之间的链接。也可以包括时间变量，以便如果用户非常迅速地选择另一个关键词，则假设前一个关键词无效并且不将其作为关键词冲浪者跟踪计数。

另外，对于每个用户，各个关键词建议器可存储其个人的关键词链接。此外，关键词建议器可以以许多不同简介类型为基础。文字结合针对不同文化，国籍，职业和年龄等的人们可能非常不同。不同关键词建议器可捕捉不同人群的关键词结合。如前面说明的，可以以与针对表3中的不同简介类型对X，Y和Z值加下标相同的方式对表10中的关键词命中加下标。

使用表格生成建议的关键词表

图21说明了可生成建议的关键词表的各种方式

一种方式是在关键词链接表696(表10)排列X的值。把排列的关键词表与来自前面相对图19的步骤646描述的正常的关键词搜索的关键词组合，如步骤730所示的建议关键词的另一种方式是把用户输入的关键词的流行表(URL的X值)与表3中其它关键词的流行表比较。表3中相似模式的X值表示这些关键词相似。例如，用户可以搜索"film reviews"，关键词建议器可以提出已更广泛地搜索的站点表"movie reviews"。这种情况下，单词"movie"和"film"之间没有直接的相似性，但他们是由URL链接的模式的相似性链接的，他们在表3中具有共性。



以与前面相对于标记的网页描述的类似的方式用安全表168和关键词链接表696确定对哪些关键词链接取样。如同用于标记号网页的决定一样，这可取决于它是否是重复的关键词(从安全表168找到)和取决于关键词使用的频率(从关键词表164找到)，以及前面讨论的考虑。

在因特网上搜索时，如所描述的显示各种不同的网页列表和网页。已描述的这些不同网页列表中的每一个的一个共同特征在于：当它们被显示时，它们表现得基本上彼此相同。如图25所示，每个不同的列表900虽然文本可能不同，在其它方面看起来是相同的。然而，其它列表902比列表900大许多倍，可包括图形内容，并且在显示给用户时表现得更加突出。该列表可包含与网页列表相同的内容，或其它内容，例如广告，图片，评论等。

在许多例子中，使用与前面相对于传统搜索引擎描述的搜索系统不同的机理，把与网页列表一起显示的内容插入显示区。使用根据如针对空间付费之类的其它标准以大尺寸显示该内容的机理。虽然用于选择该内容的系统工作，很难保持显示内容的踪迹，特别是当如果频繁改变该内容时更是如此。因此，现在讨论的本发明的另一个方面是用于跟踪改变的内容，并在显示其内容时允许内容提供者动态选择的一种系统。如图22所示，可根据所示的步骤762中由观众输入的关键词或简介类型向观众显示这种动态可选择的内容。在随后的内容选择器步骤764中，考虑当天的时间，并在如图25所示的

关键词	一个月的累积命中(H)	发出内容的数量(N)	内容提供者1(A1)	内容提供者2(A2)	总和(T)	(P)
Book						
Fish						

表12是关键词内容数据集

该表还可包括准备给予内容提供者的最大内容值M。对可能尝试用与特定关键词相关联的网页列表显示内容902的内容提供者的数量没有限制。

每个国家或地区可能有分离表12，以便每个国家或地区，每个关键词的内容值可以不同。另外，在每个国家或地区中不同的时间周期可以有不同的内容值。

内容902的提供者可通过识别具有目标读者的每个关键词来以关键词和读者为目标，例如，与单词橄榄球相关联的命中数量可能分解成搜索单词橄榄球的不同简介类型。搜索橄榄球的累积数量可能是6000人，分成21岁以下年龄组520人，21-50岁年龄组4000人，和50岁以上年龄组520人。因此，对于关键词搜索内的这些小类中的每一个可以有不同内容值。

除了关键词数据集804外，可以有列类型的不同简介类型806的数据集。它包含每种简介类型的相同项目，代替相对于图23的关键词内容数据表804如上所述的关键词。

简介类型	一个月的累积命中(H)	发出内容的数量(N)	内容提供者1(A1)	内容提供者2(A2)	总和(T)	(P)
男性						
女性						
职业等						
未定义简介						

表13是个人简介内容表

与不同的关键词相反，表13将内容902的内容值确定为特定的读者群，并允许以特定读者为目标。

本发明的范围还包括表13中的组合简介类型，例如，男性，职业或新西兰，女性。组合的简介的内容值与单个简介的内容值不同。确定内容值和选择内容902中涉及的机理是相同的，并在下面进一步说明。

图23的内容提供者数据表812在下面按表14更详细地说明并包含有关内容提供者的诸如姓名，地址，广告商之类的信息，内容信息，例如内容902将要使用的Bitmap(HTML或Java小应用程序或类似内容)，以及识别内容902的每个不同项目的唯一编号。

姓名	地址等	内容信息	每个内容的唯一号码
E.g. John			内容号码
			内容号码

表14

该表也可存储内容提供者的详情，例如密码，付费详情(例如信用卡号码和授权)，内容传送(已向用户发送内容的次数)等。

上面提到的内容表的数据集被占据如下。对于关键词内容数据表804

- H，从表1(800)直接取出一个时间周期对特定关键词的命中累积数。

- N是发出与该关键词相关联的内容的次数。每当向用户显示内容902的与该关键词810具体相关联的项目时将其递增。

- 由每个关键词的内容提供者选择A802的值。内容提供者还可输入最大值M，在该最大值范围不再选择将要发出的关键词。

- T是每个关键词的总和并且是A的和

P是由投票或价格针对每个关键词确定的内容值并且是 T/N

传播个人简介内容数据

·H是从表1(索引W的和)直接取出每个简介类型和该信息的命中累积数。

·N是发出与该个人简介相关联的内容902的项目的数量。每当发出内容902的与该简介类型810具体相关联的项目时将其递增。

·通过类似于投标的输入处理对每个简介类型安排A808的值。内容提供者还可输入最大值M,准备对他们付费或投票,视情况而定。

·T是每个简介类型的总和并且是A的和

P是每个简介的内容值并且是 T/N

传播内容提供者的详情表

由内容提供者电输入内容提供者的大部分详情812。每当发出内容提供者的内容902时,该事件也记录在内容提供者的详情表812中。还要记录点击(820, 822, 824, 826, 828)次数和在付费或投票方面内容902的费用。这是形成此后传送到内容提供者的电子票据或制表的基础。

如何使用数据集选择发出到用户的内容

在下面参考图24的讨论中,假设利用每组网页搜索结果900仅发送内容902的一个标识。如果利用每组网页结果发送多组内容,则应用相同的算法。

在步骤852向搜索引擎提交关键词和简介类型。从表中的P值找到关键词内容数据表804,个人简介内容数据表806,内容902的值中的每一个。在步骤862确定的关键词或简介类型的P的最高值确定与网页列表900一起发送的内容(关键词或简介类型)的类型。可能没有针对该关键词的具体值并且用户可能不使用具体的简介类型。这种情况下,使用未赋值的内容项目的值(来自没有简介的用户的表130。下面讨论发出具体内容项目902的选择。从表14,内容提供者详情表814取出该内容项目的详情(其图形,文本,相关的程序等),并在步骤868发射到用户。为每个内容提供者发射的内容项目902的详情同样以规定的间隔传送到内容提供者,如步骤870所示。

确定它是否是发射的关键词或简介内容

所发射的内容902的类型取决于它是否是基于关键词的内容或基于简介选项的内容。例如，来自美国的男性可搜索钓鱼。可应用到该搜索的值是：关键词=钓鱼，简介=男性，简介=US，简介=US，男性。在决定显示哪个内容时，该系统比较所有可能性的内容(关键词，简介类型的组合)的值，并发出如步骤862确定的具有最可能的值的内容。例如，21岁以下的男性可能使用关键词"Rugby"搜索，并且与Rugby相关联的内容的值是每个画面0.1，而以21岁以下的男性为目标的每个画面的值是0.2，因此，显示以21岁以下的男性为目标的内容而不是橄榄球内容。指出发射内容项目902的累积次数频率(N)不同于目标区的总累积频率(H)是很重要的。在该例中，搜索'rugby'和'21岁以下男性'的次数的累积频率(H)都加1(通过表1)。然而，仅对'21岁以下男性'表增加显示内容902的项目的次数(这是用来确定每个单元图象的内容值的图形)。

确定发射哪项具体内容

下面的例子给出如何选择与该关键词相关联的内容。该过程与和简介类型相关联的内容相同。

关键词	一个月的累积命中(H)	发出内容项目的数量(N)	内容提供者1(A1)	内容提供者2(A2)	总和(T)	(P)
Book	134	134		10	10	0.050
Fish	52	80	5	5	10	0.52

对于关键词"book"，每当搜索该关键词时，显示内容提供者2的内容902，由于他们仅是与该关键词相关联的内容提供者。然而，对于关键词"fish"，内容提供者1和2将使其内容发出相同次数。在换算成将要使用的等级的系统中，有很多对不同关键词和简介类型报价的内容提供者。

计算内容值

如果有一个针对关键词"book"的新内容提供者, 该内容的值为例如每月\$5, 将该值改变成0.075, 这表明与单词"book"相关联的总和是\$15. 因此, 内容提供者会得到66% (10/15)的发送时间, 新内容提供者显示33%的时间. 发送内容提供者的内容的比例是 A/T .

内容提供者如何使用数据表

当对内容902报价时, 内容提供者从表12&13选择以其内容为目标的关键词或简介. 搜索引擎自动表示对前一个时间周期(H)执行该搜索的次数, 发出与该选择相关联的内容项目的次数(N)和内容P的值.

此后, 新内容提供者输入选择系数A, 然后该系统可根据新的总报价(T)立即计算新值(P). 广告商也可告诉每月的画面数量, 他们很可能得到其报价($N*(A/T)$). 实时计算这些变化以便向新内容提供者给出其报价将如何影响该值和他们将接收其报价的画面的指示. 如果值和画面的数量与适合于该广告商, 他们可进行选择以提交其作为针对定义的周期, 例如一天, 一星期, 一个月的报价. 最好不对其它内容提供者的详情进行公开. 内容提供者也可输入最大值M, 他们可与其内容分离. 这样向内容提供者提供了一些安全性, 以防止如果该值改变时付费太多. 如果该值太高, 内容提供者的报价则可脱离该表(如果P大于M, 则不认为A是对该特定内容提供者的报价). 如果该值再次下降, 报价将回到表上, 从而起到稳定机理的作用. 在优选实施例中, 如果其内容902因其值M的限度而脱离该表, 则用电子邮件通知内容提供者.

如图24的内容提供者详情表812所示, 例如, 内容提供者通过搜索引擎的所有者而有一个帐户, 并在定期的基础上从帐户详情自动计算针对其内容记入其帐户的借贷的过程. 还将画面数量, 每个画面的费用, 点击次数和针对每个内容提供者的每次点击的费用的电子语句传送给每个内容提供者, 因为该信息也存储在内容提供者详情表812(表14)中.

在优选实施例中，可根据关键词链接表识别相似的关键词串。识别关键词串的原因是为了能够以单词组而不仅是单个单词标记内容902。关键词串"car"可包括数百或数千个与单词"car"有联系的单词(例如，两用车，汽车，带蓬车)。统计串技术被用来定义关键词串的规模和频率。这使其与一个编辑决定内容提供者的作为目标的关键词串相比是一个更高度自动的过程。

可使用相同系统设定关键词串的值。虽然以这种方式分组的单词会招致管理费用的增加，尽管如此，他在计算上相似并且仅在超过对一个关键词的命中的特定等级的情况下启动。

仅搜索用户的内容也可有目的地选择，以便仅搜索与关键词相关联的内容提供者。这种情况下，搜索结果以表12中的A的值为基础。付费最多的内容提供者在列表的顶部。

关键词建议器也可帮助内容提供者选择他们喜欢显示的关键词或关键词组。

控制搜索引擎系统

存在许多可改变根据本发明的搜索引擎排列网页的方式的参数。这些因素(下面详细说明)是：

· 历史系数

它确定了如本文前面描述的现有流行表(流行命中表)的下降速度。该数值在1和0之间。高历史系数将使其很难改变现有的流行表。作为例子，如果搜索特定特定关键词的速度迅速增加，该历史系数应降低，以使出现的网页能够在流行表中上升。

从累积的冲浪者跟踪更新表3的频率

这是用关于用户活动(即冲浪者跟踪)的信息更新流行表的频率量度，例如，根据特定关键词搜索的流行性改变速度可一天或甚至一个月测量一次。

· 取样频率

这是对用户如何进行搜索的信息取样的频率。如果它是一个常用关键词，则不必监视每个搜索。可能仅需要监视所有搜索的一个

百分比，以便准确地确定网页流行性。

- 缺省搜索表的组合(来自新网页表，高跳跃和流行表等的结构的混合)

如果需要反映搜索结果在一段时间演变方式，可改变作为缺省给予用户的网页混合。

- 内容'命中系数'

"内容命中系数"是给予作为关键词的命中记录的有关的内容的命中的加权的量度。该缺省设定认为有关内容的命中与来自网页表的命中相同。可根据内容的价格设定内容命中值大于或小于1，例如，对于可变关键词可能需要增加"内容命中系数"，也可减小该能力以便滥言(spam)那些大批的可变关键词。内容系数越高，对滥言的阻力越高，该搜索结果更多地取决于价格而不是流行性。

- 内容报价的时间周期

内容提供者针对特定时间周期，例如一个月报出某个数值。该时间周期可根据价格变化率而不同。如果该价格变化快或非常稳定，可相应地分别缩短或延长该时间周期。

- 每次网页提交的关键词数量

可改变该数值以影响该系统从新网页提交了解多个内容

如果网页的提交日期与现有的网页提交太近，则是不能接受的。可根据上面的任何因素改变该时间长度

- 每天，每人(IP地址或用户ID)认为有效命中的搜索数值。可改变该数值以减少滥言的可能性

- 更新安全表前的时间长度

限制滥用的安全表指出了关键词与用户识别的IP地址之间的链接。可改变刷新该表之间的时间长度以使其很难滥言该系统。

对这些系数的设定可针对不同的关键词或人群而不同，取决于：

- 进行搜索的频率

- 搜索频率的变化速度

科学刊物

用本发明搜索科学数据库(在线论文, 期刊等)将显著减少察看含糊, 或深奥的领域而仅找到无关信息所花费的时间。对于这种用途, 有效命中的标准通常体现了上述延伸的时间特性, 以建立信息来源的有用性。通过分成不同等级或用户类型, 例如博士, 教授, 硕士等可提高使用本发明对学术/科学论文的查阅和参考。例如, 这样能使用户看到什么信息来源于找到的感兴趣的特定领域中的著名权威。它还允许论文的作者了解其出版物被如何频繁访问, 并可能进一步指出该论文在以后的论文中的哪部分和如何频繁地被用作参考。用户可能必须正式注册不同的组织以获得查阅能力的等级。用户也可选择查阅其搜索的等级。

在线帮助

目前对在线帮助和支持存在着相当大的全球需求, 特别是对计算机/软件应用。由于软件开发者获得对特定查询的类型和频率的直接反馈, 使用本发明会明显减缓这种需求, 同时用户收到了以前用户积累的优点。不同的简介类型能够以适当的形式为用户, 例如新手, 专家等提供答案。关键词建议器可以例如建议用最可能产生满意答复的关键词(问题)进行搜索。对每个问题可能有一批答案, 该系统学会将其集中在最好的答案上。

问与答服务

通过运行本发明可形成目前的在线问/答节目, 因此能够回答根据前面的问题和将要提出的类似问题重复询问的问题。

对因特网其它部分的内容优化

可以使用相同的内容报价机理确定因特网上任何位置的内容的价格, 而不仅仅是如上面识别的网页表。在该实施例中, 内容提供者对一般的内容空间报价以便自动设定价格。

可使用来自搜索引擎的简介类型信息作为护照, 以便因特网上的其它广告能够以不同的读者为目标。也可由网页开发者使用该简介类型信息, 以便为不同的用户组定制其网页。

人群选配服务

在另一个实施例中，可根据本发明的系统用作约会服务和/或通过进行统计分析以比较用户群的各种喜好(表6)来选配具有相似喜好的人群的方法。在该实施例中，最好使用标准相关系数规范和相互比较各种过去的喜好表。在与其它用户比较时，将给出大量其喜好如何相似的指示。

也可使用同一实施例从过去的喜好表找出有关类似人群的信息。例如，人们可请求给出在新西兰对生态经济感兴趣的人的姓名并可对个人喜好表进行搜索。然而，该实施例通常包括提供被识别人给出其信息的许可的密码/许可指示器，该许可仅在例如特定环境中给出，该环境限制在具有职权和表示该职权的密码的等级的搜索者，或用特定特征识别其自身的人们。

虽然已结合目前被认为是最实用和优选的实施例描述了本发明，应该理解，本发明不限于所公开的实施例。例如，如下面的权利要求书所述的，可单独或组合使用上述每一个特性，上述未描述的其它特性其本身明显具有专利性。因此，本发明意图在于覆盖所附权利要求的精神和范围内包括各种改进和等效安排。

说明书附图

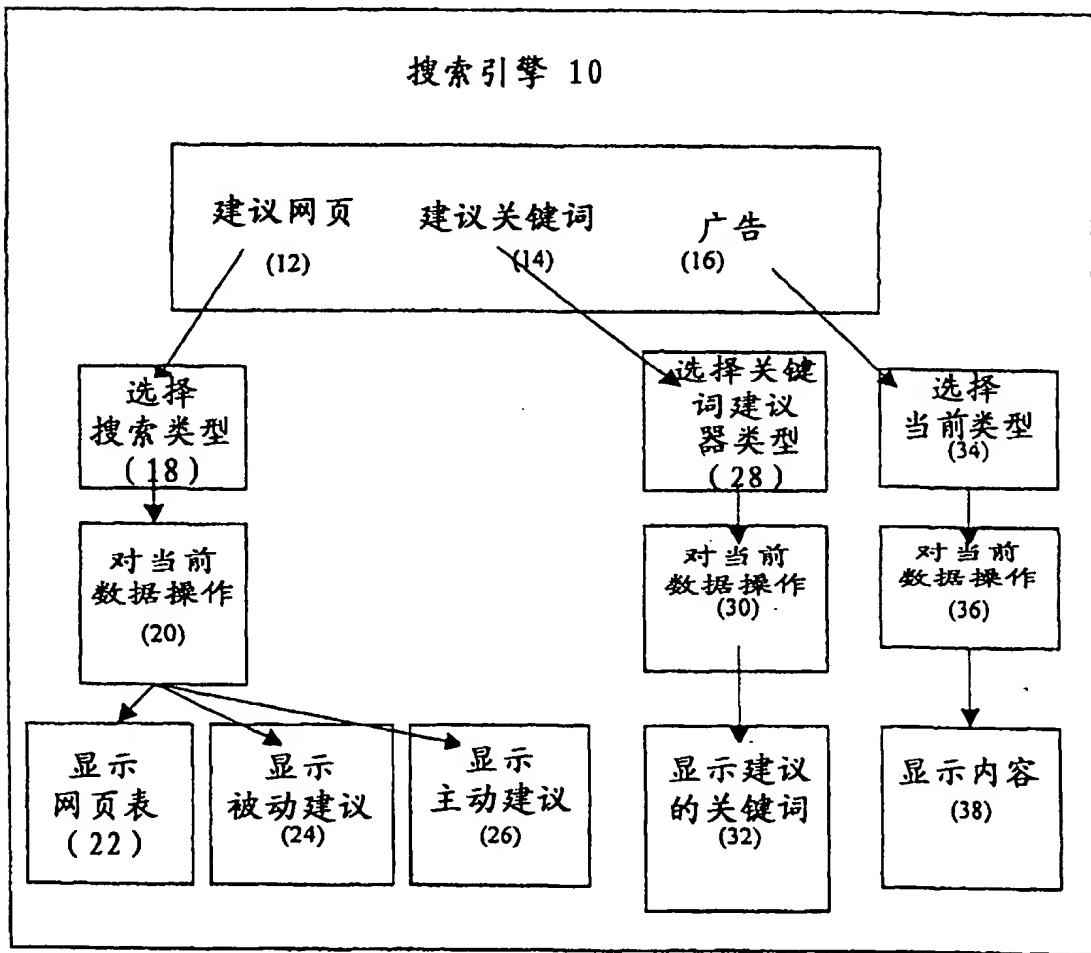


图1A

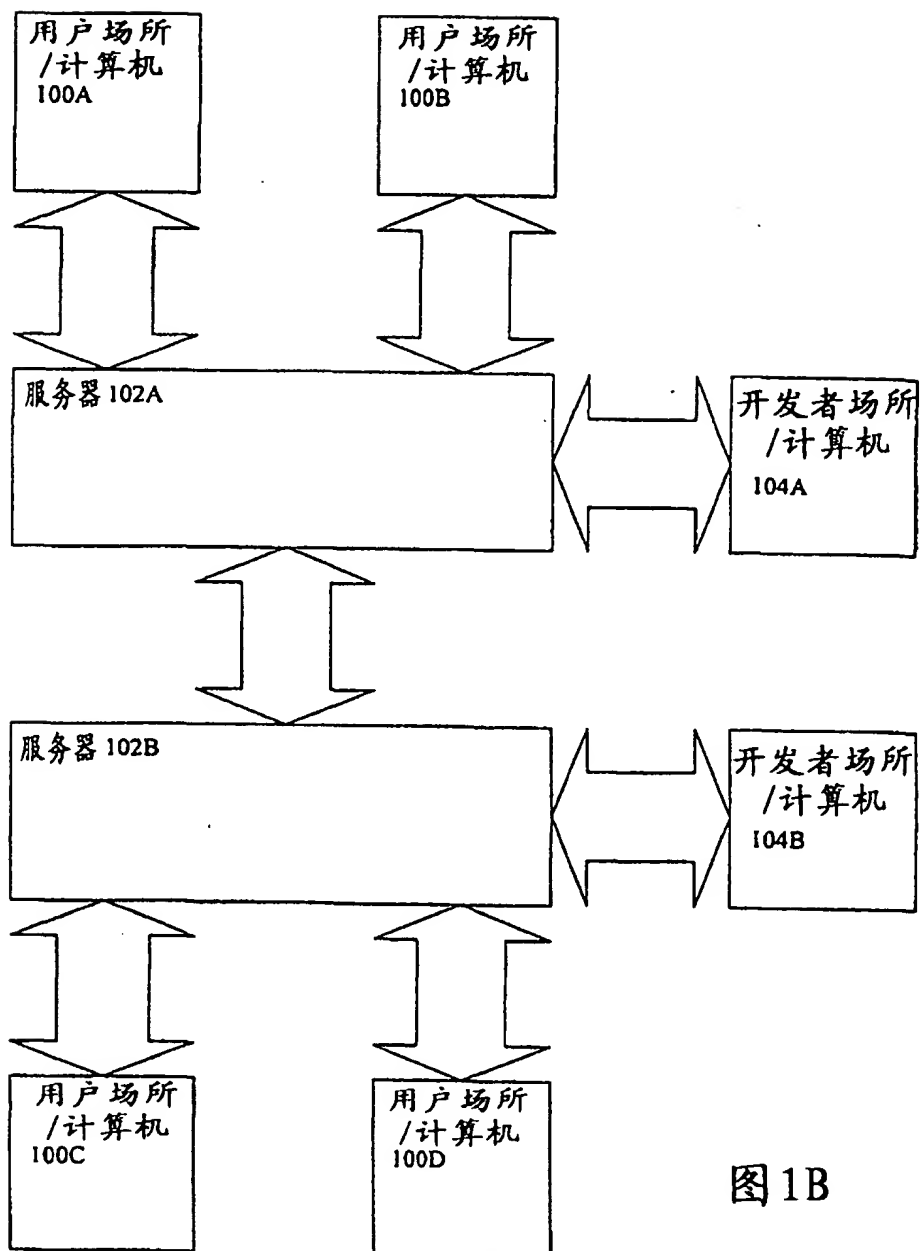


图 1B

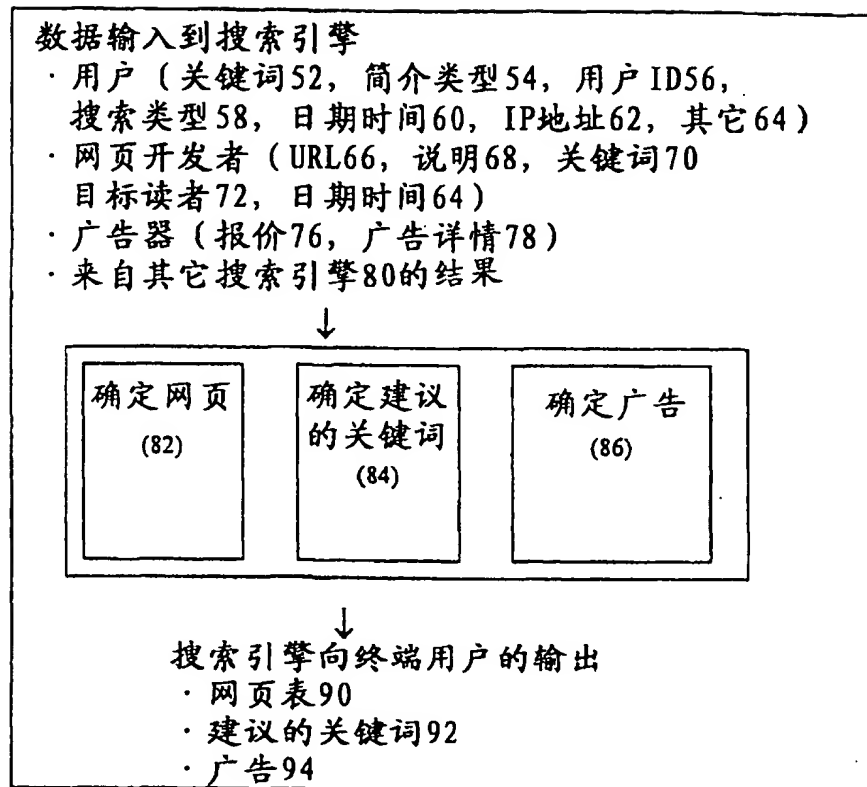


图 2

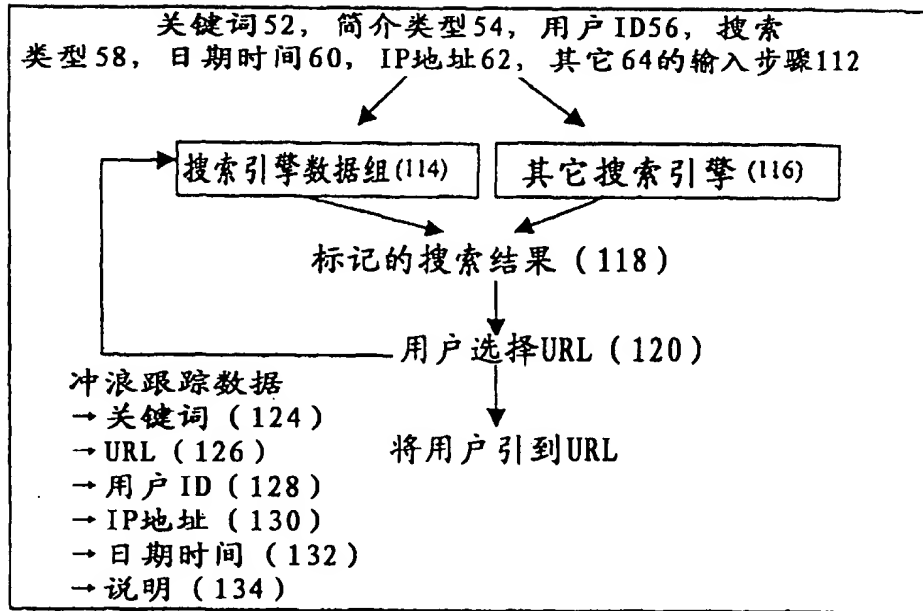


图 3A

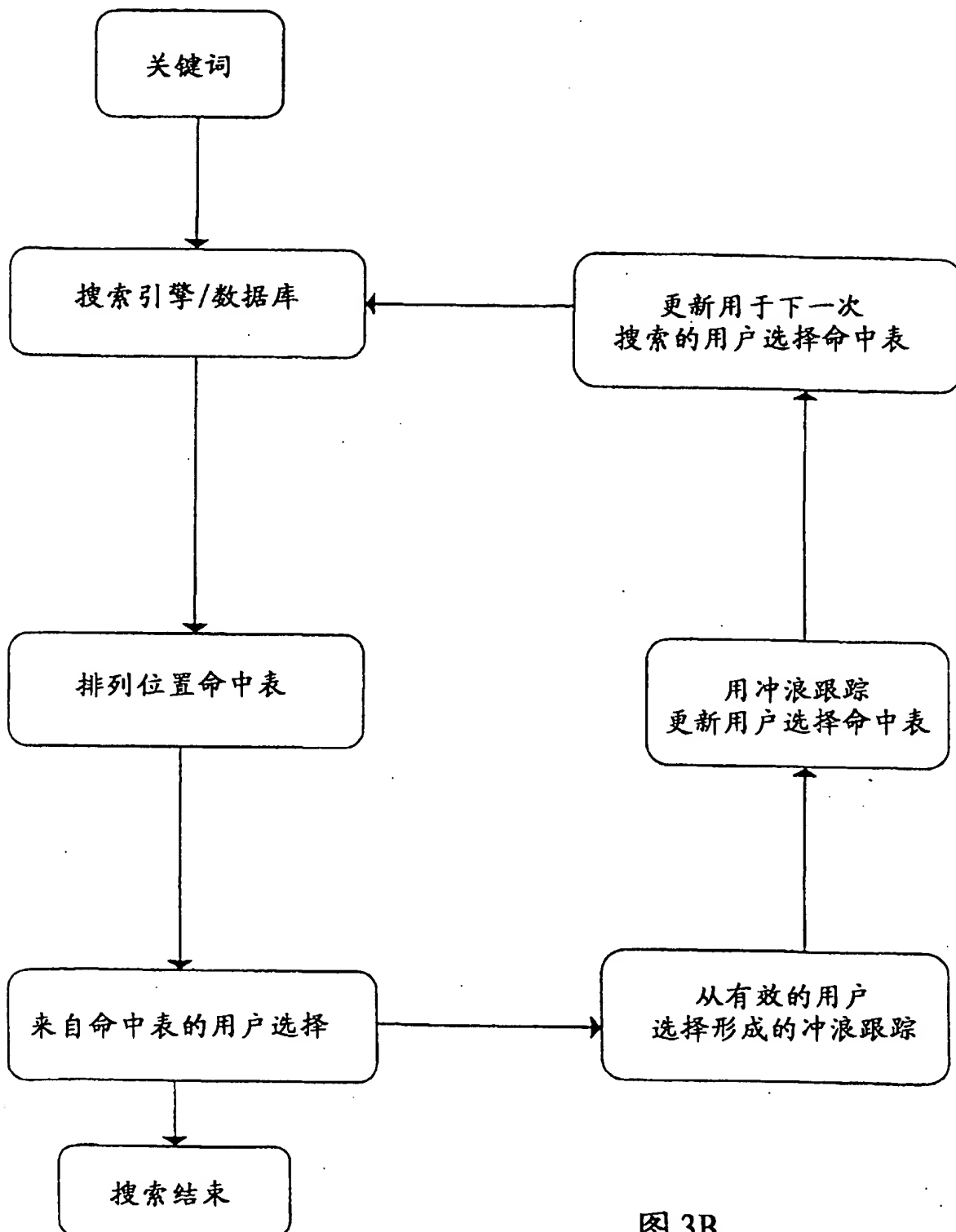


图 3B

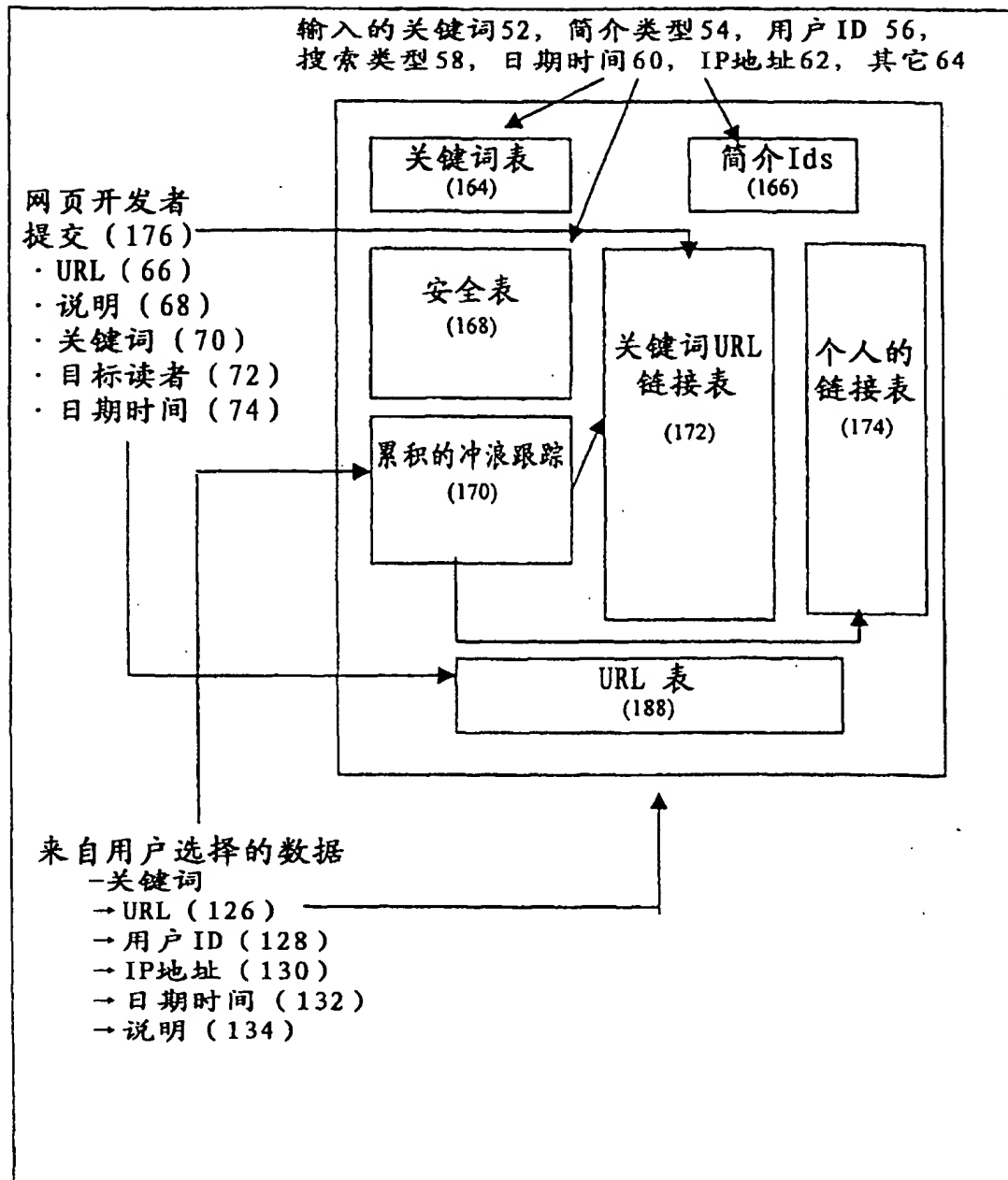


图4

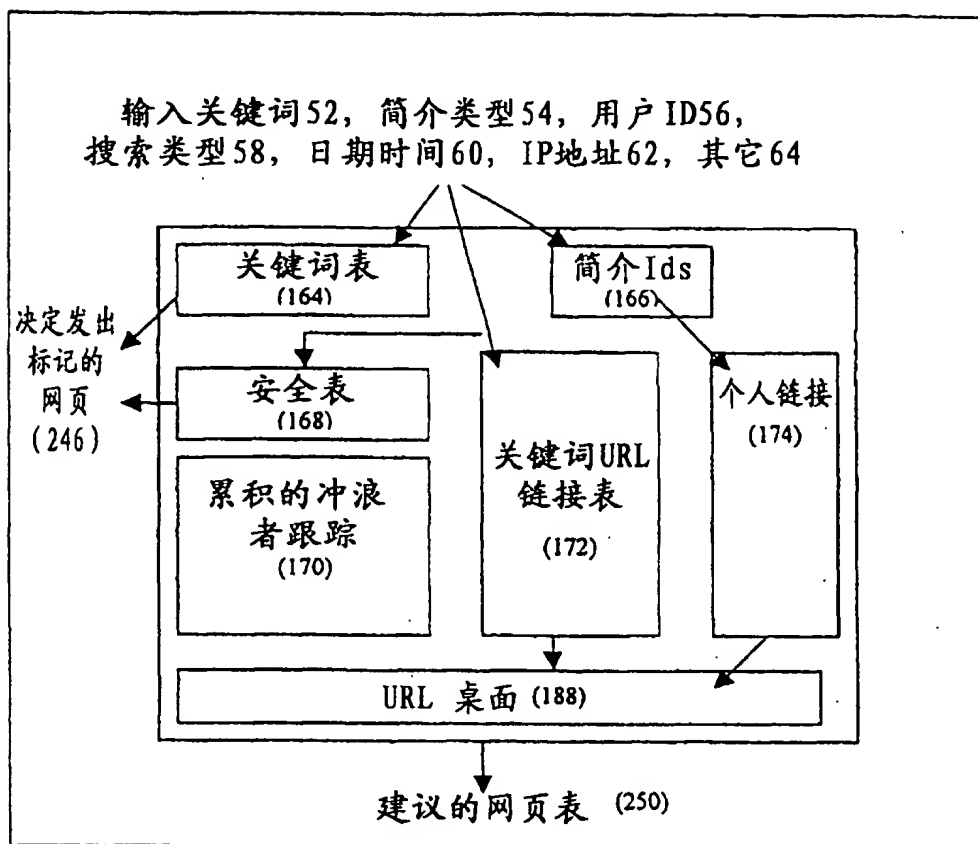


图5

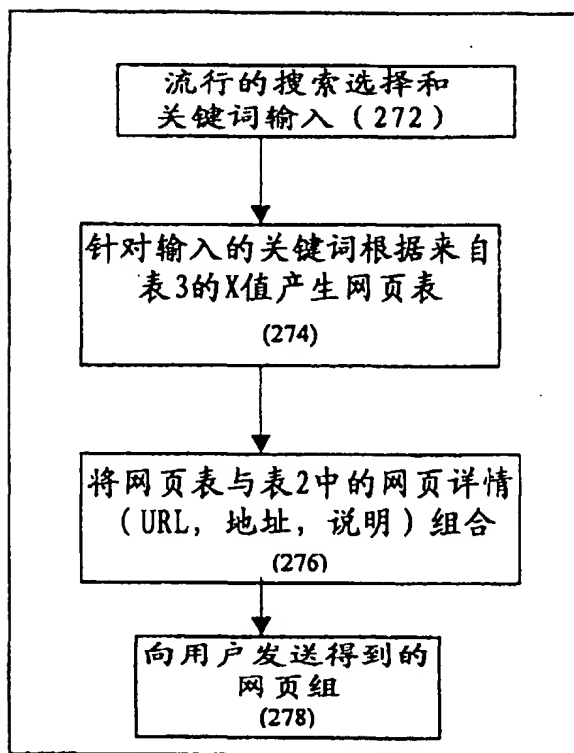


图 6

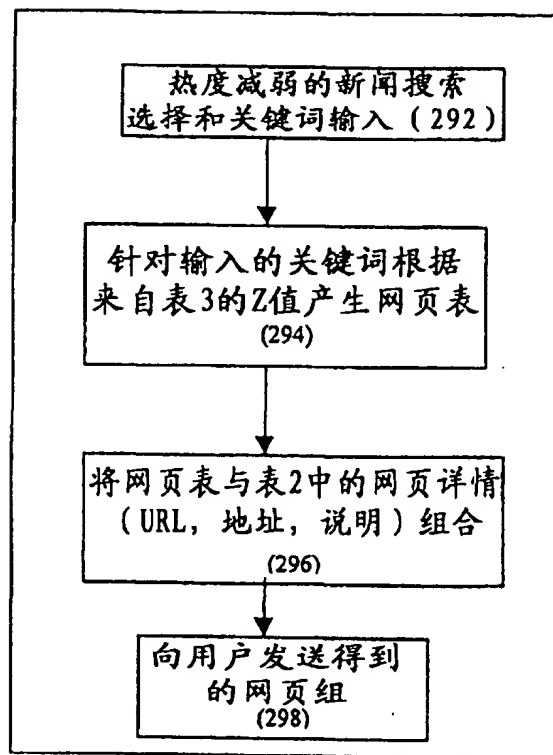


图7

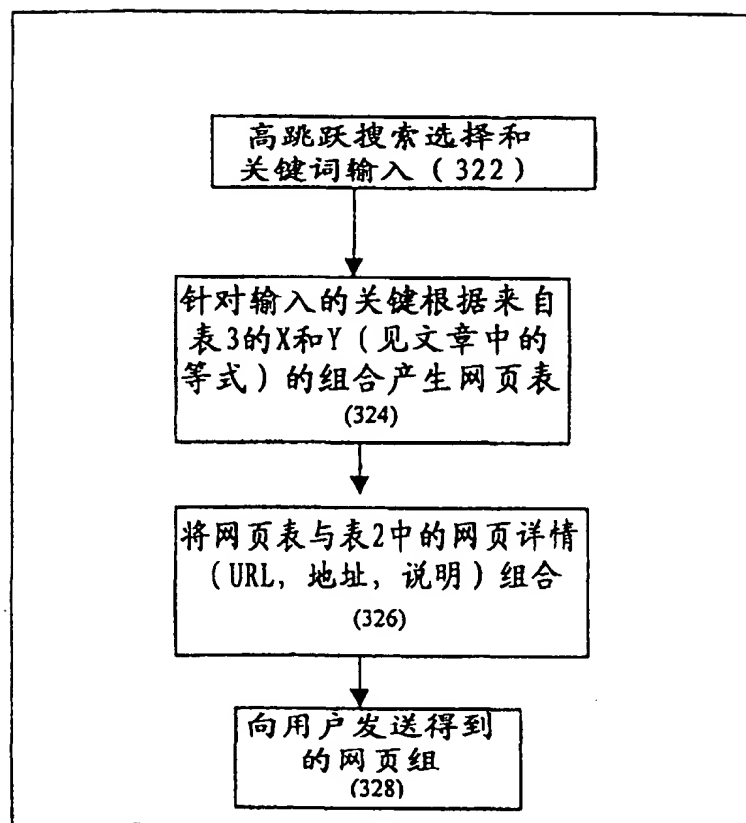


图8

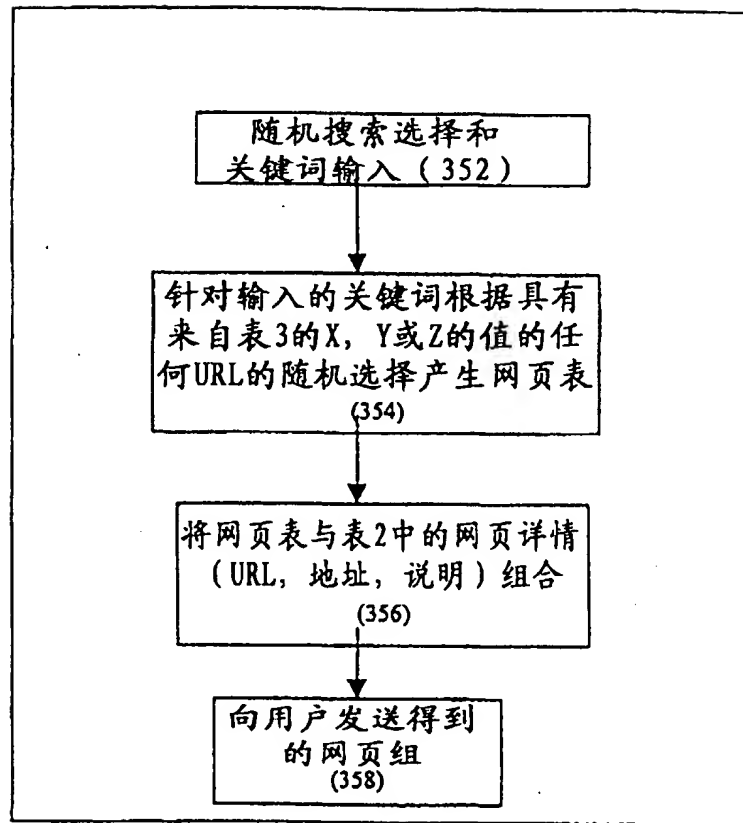


图9

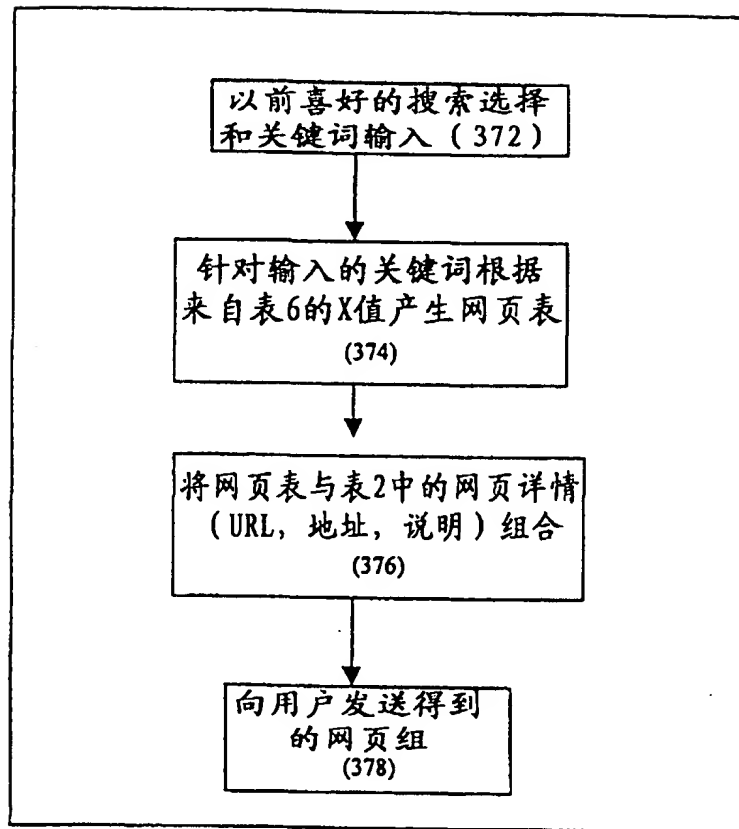


图10

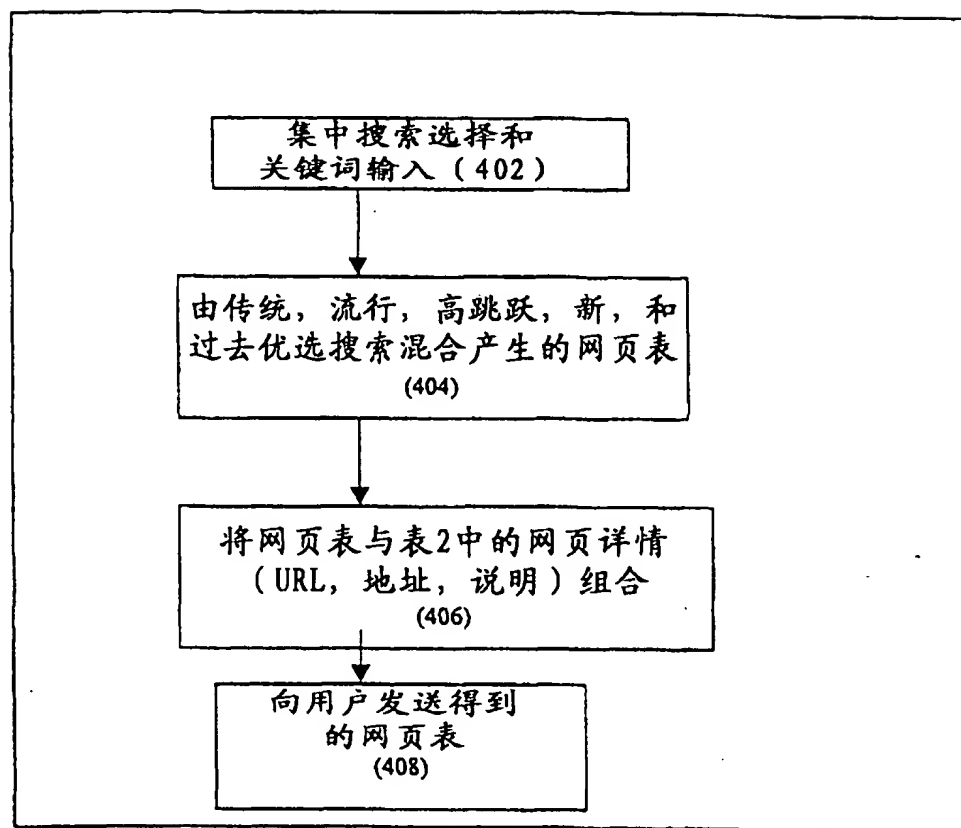


图 11

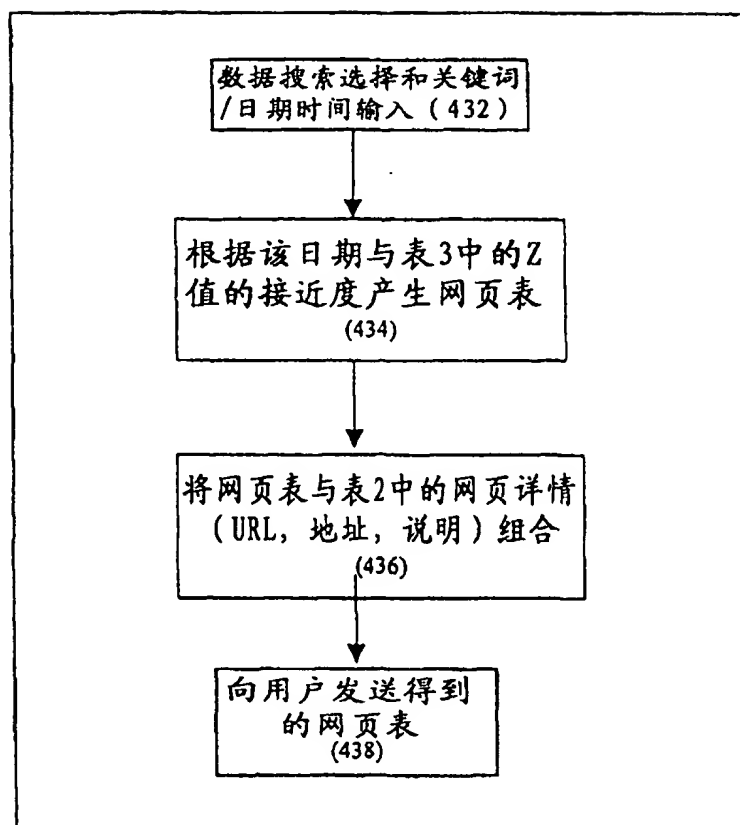


图 12

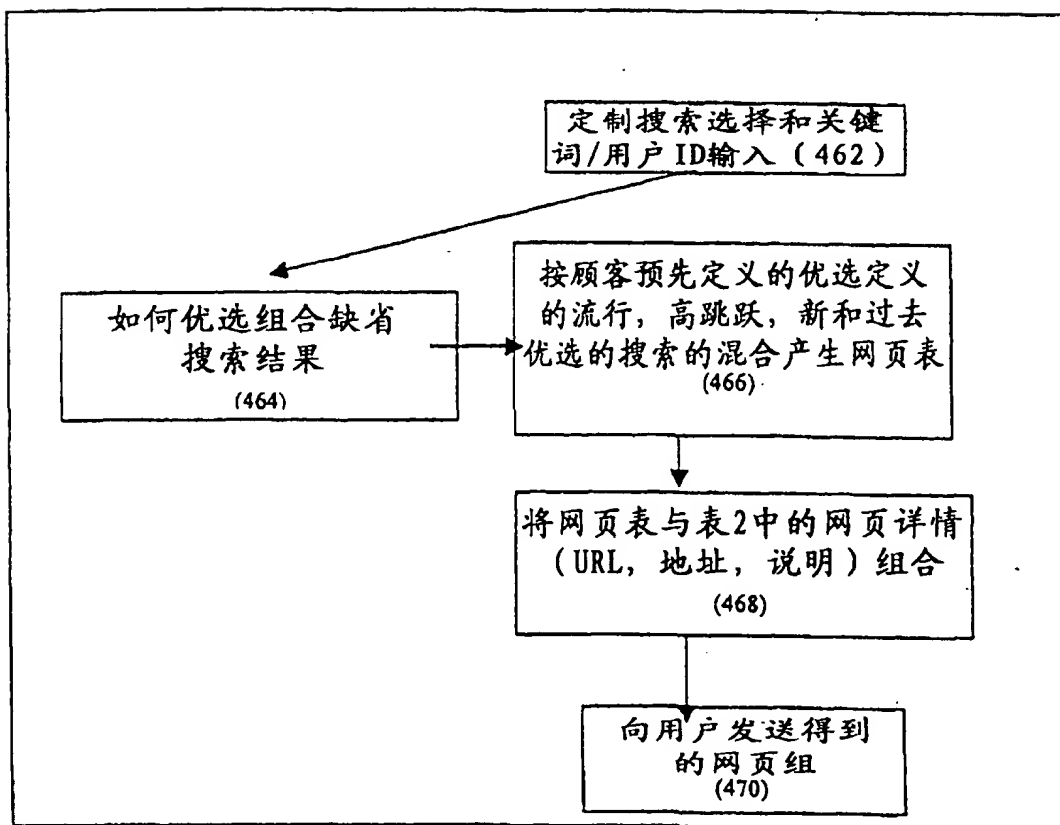


图 13

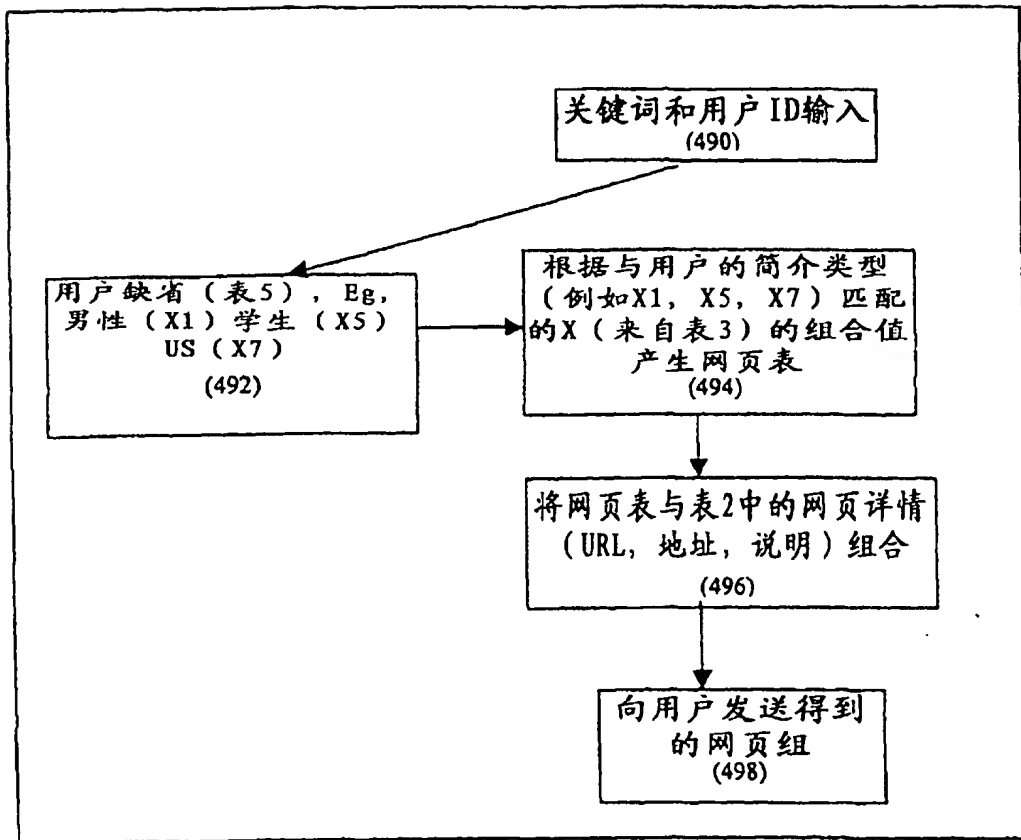


图 14

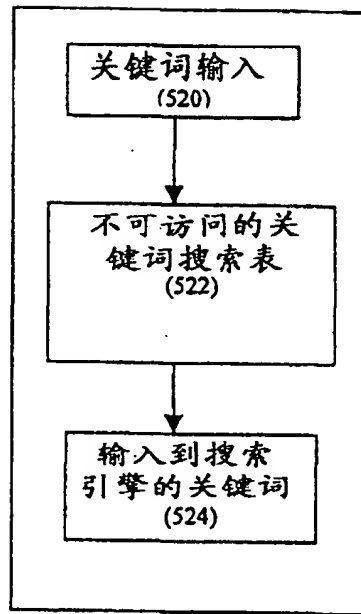


图 15

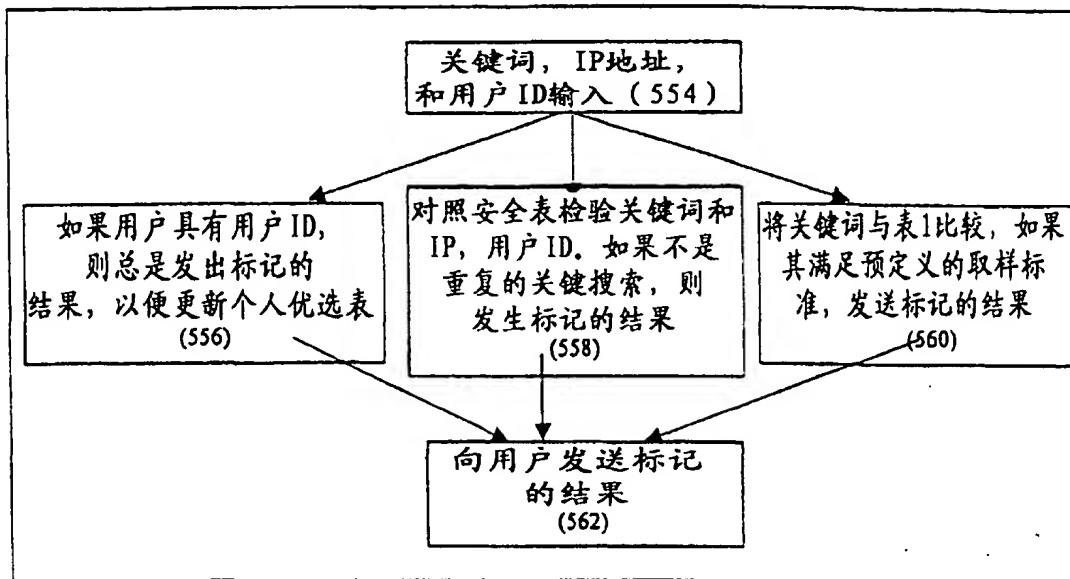


图 16

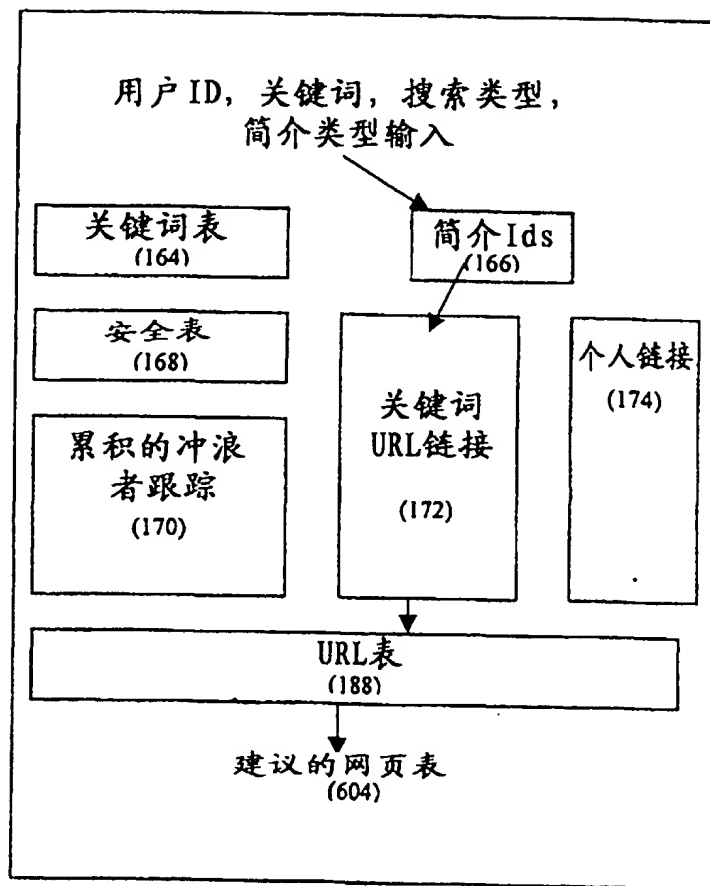


图 17

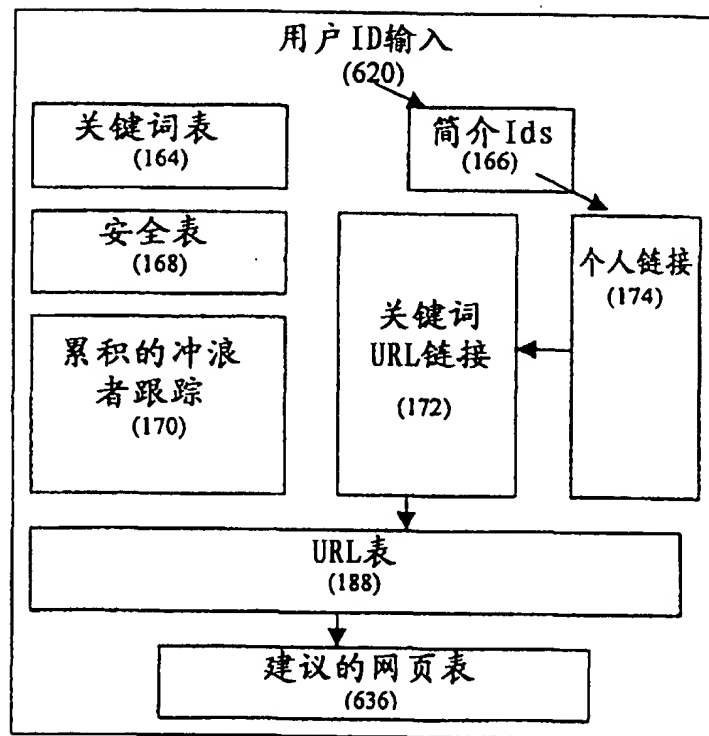


图 18

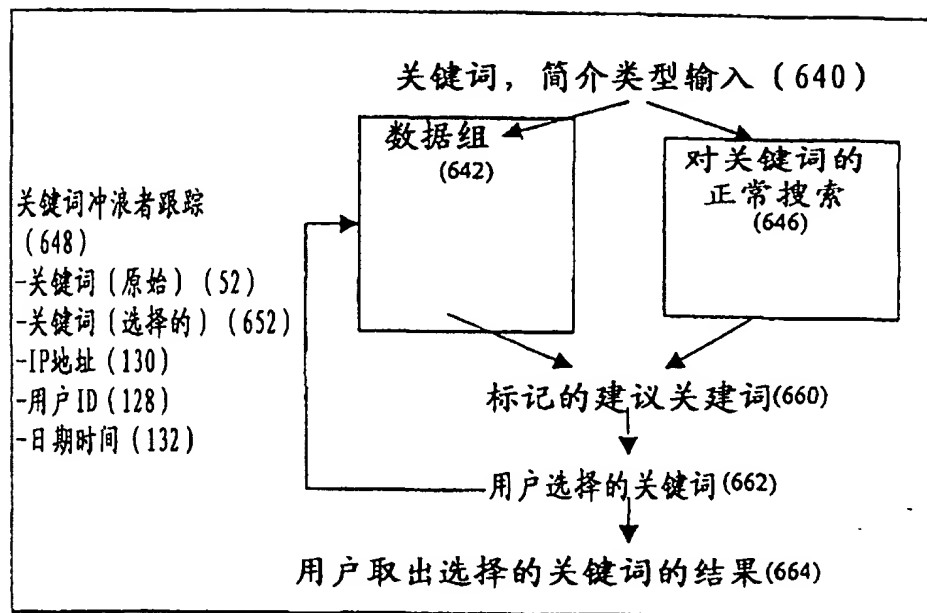


图 19

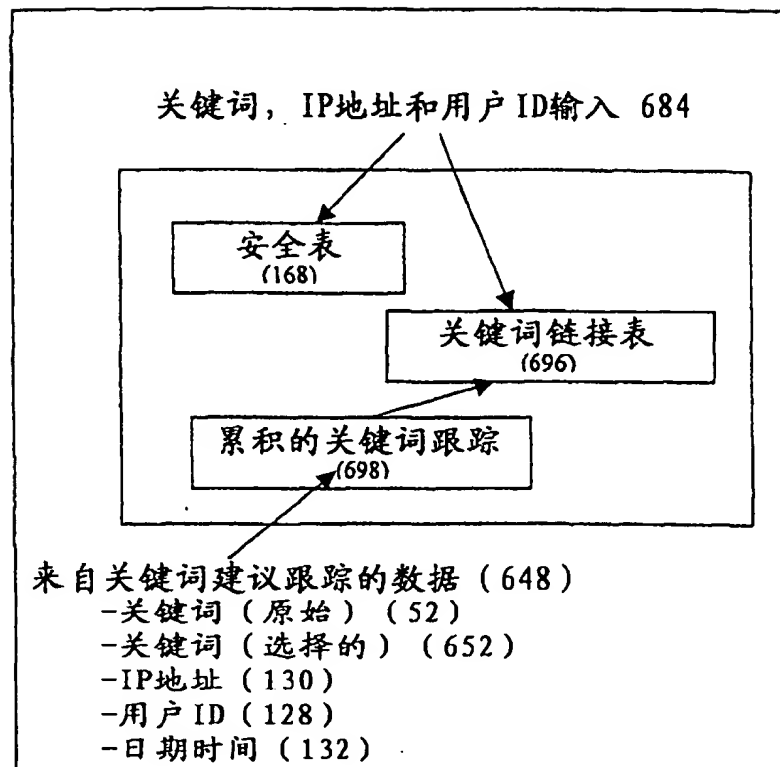


图20

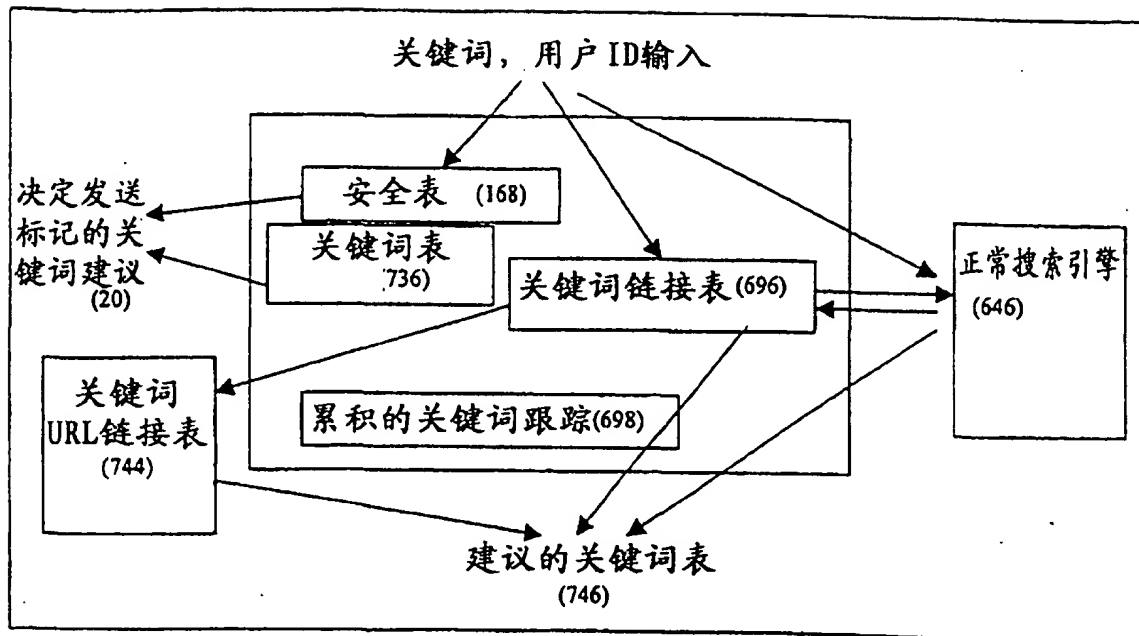


图 21

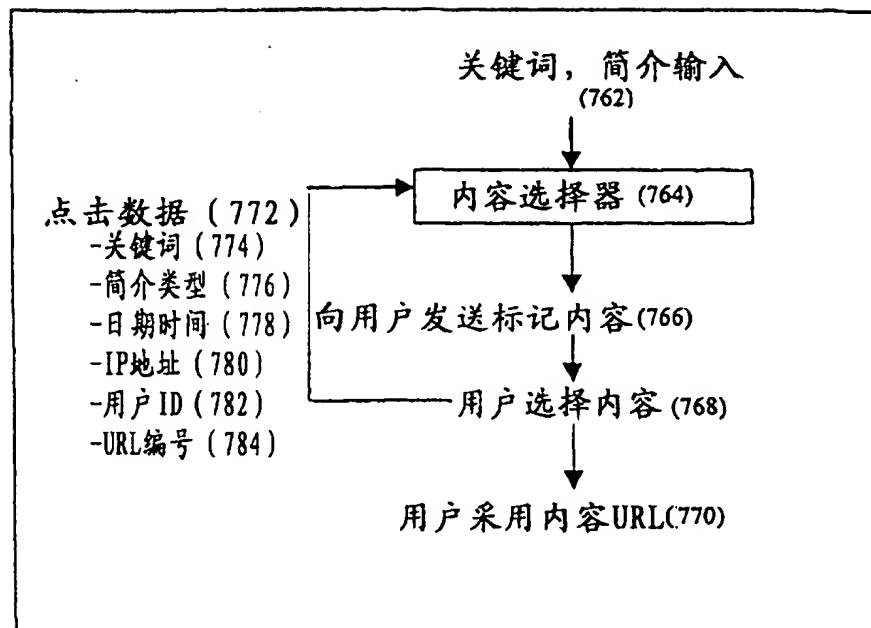


图22

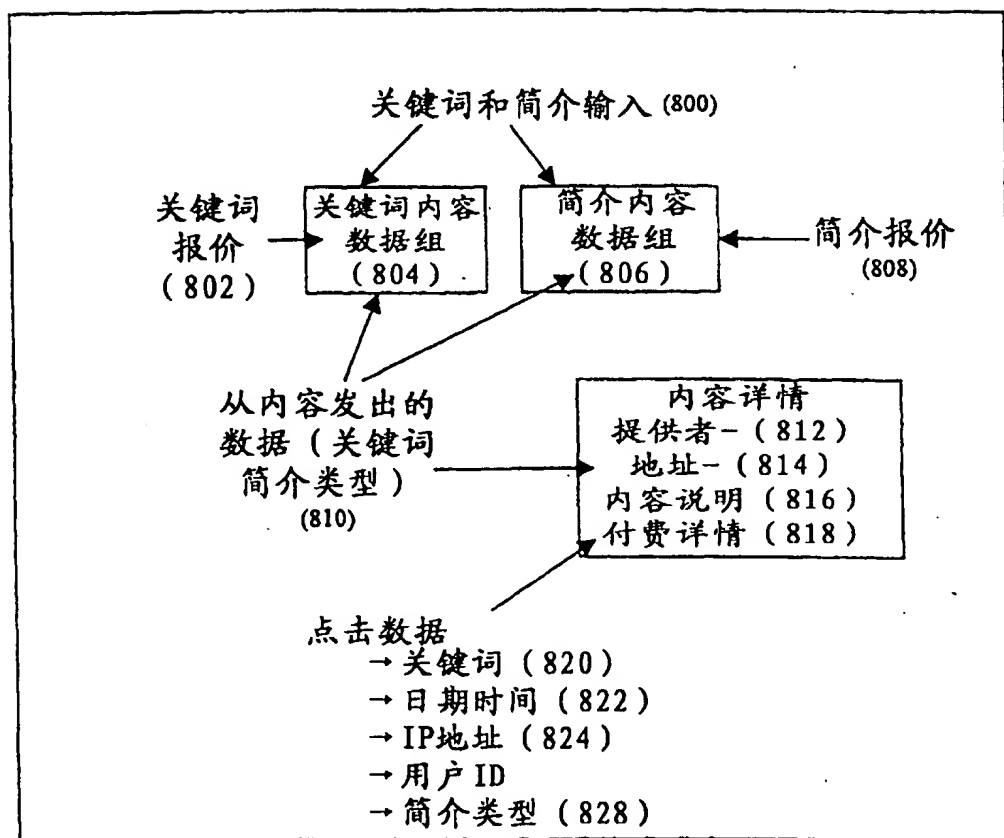


图 23

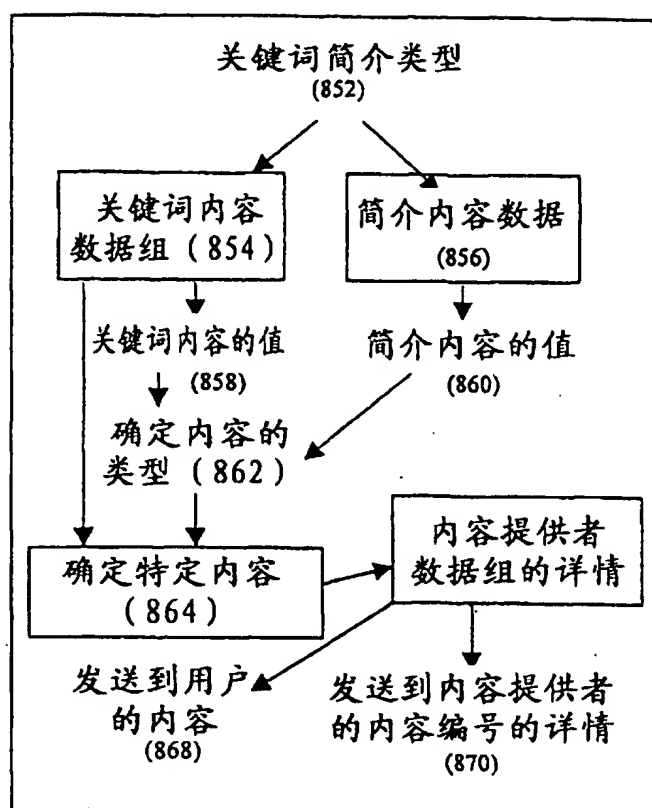


图 24

27/27

	内容 902	
	网页表 900	
	网页表 900	
	网页表 900	

图 25

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.